

Koustuv Saha, Ayse E. Bayraktaroglu, Andrew T. Campbell, Nitesh V. Chawla, Munmun De Choudhury, Sidney K. D'Mello, Anind K. Dey, Ge Gao, Julie M. Gregg, Krithika Jagannath, Gloria Mark, Gonzalo J. Martinez, Stephen M. Mattingly, Edward Moskal, Anusha Sirigiri, Aaron Striegel, and Dong Whi Yoo. 2019. Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19). ACM, New York, NY, USA, Paper CS12, 8 pages. DOI: <https://doi.org/10.1145/3290607.3299065>

Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing

Saha, K., Bayraktaraglu, A. E., Campbell, A. T., Chawla, N. V., De Choudhury, M., D'Mello, S. K., Dey, A. K., Gao, G., Gregg, J. M., Jagannath, K., Mark, G., Martinez, G. J., Mattingly, S. M., Moskal, E., Sirigiri, A., Striegel, A., & Yoo, D. W.

KOUSTUV SAHA, GEORGIA TECH



Sensing Human Behavior



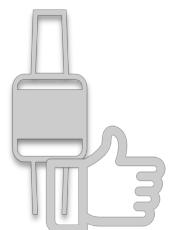
Survey Instruments

- Self-Report Questionnaires



Active Sensing

- Ecological Momentary Assessments (EMAs)



Passive Sensing

- Smartphones and Wearables
- Social Media

The Tesserae Project



By leveraging passive sensors, this study aims at proactively identifying changes in an individual that may impact their wellbeing and job performance



Wearable



Smartphone



BT Beacon



Social Media



Surveys

The Tesserae Project



By leveraging passive sensors, this study aims at proactively identifying changes in an individual that may impact their wellbeing and job performance



Wearable



Smartphone



BT Beacon



Social Media



Surveys

Social Media as a Passive Sensor

- ▶ Naturalistic setting
- ▶ Unobtrusive access
- ▶ Longitudinal and Extended Periods (beyond study period)
- ▶ Verbal and Behavioral

Not always easy to collect



This case study...

...introduces an infrastructural framework to illustrate the feasibility of collecting social media data at scale. This is in the context of an ongoing multimodal sensing study of workplace performance

Social Media Data Collection Infrastructure

- ❖ Facebook, LinkedIn, Instagram, Twitter, GMail, Calendar
- ❖ Open Authentication (OAuth)
- ❖ Social Media Authorization per platform
- ❖ Python Web application using Django framework
- ❖ Models-Views-Controller (MVC Architecture)
- ❖ Hosted on a secure and encrypted server

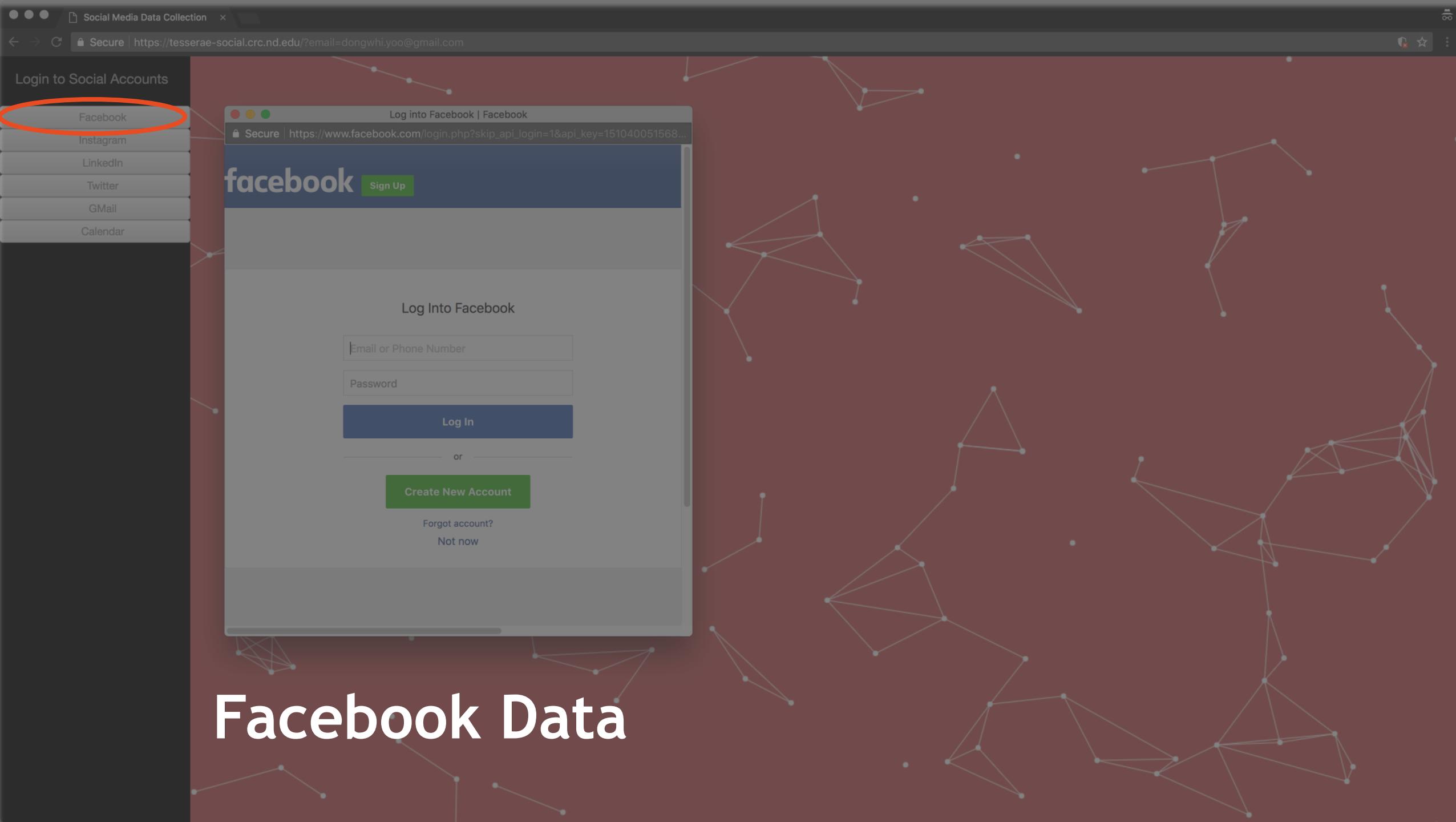


Tesserae

Tackling Developmental Challenges

- ❖ Continuous development and testing with use-cases and automated scripts for debugging
- ❖ API changes during ongoing data collection
- ❖ Cambridge Analytica breach, and more comprehensive application approval

The screenshot shows the header of The Guardian website. On the left, there are 'Sign in' and 'Contribute' buttons. The main title 'The Guardian' is prominently displayed. Below the title, a navigation bar includes links for 'News', 'Opinion', 'Sport', 'Culture', and 'Lifestyle'. To the right of these links is a yellow circular icon with three horizontal lines. The main headline reads: 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach'.



Facebook Dataset: Participant Authorization

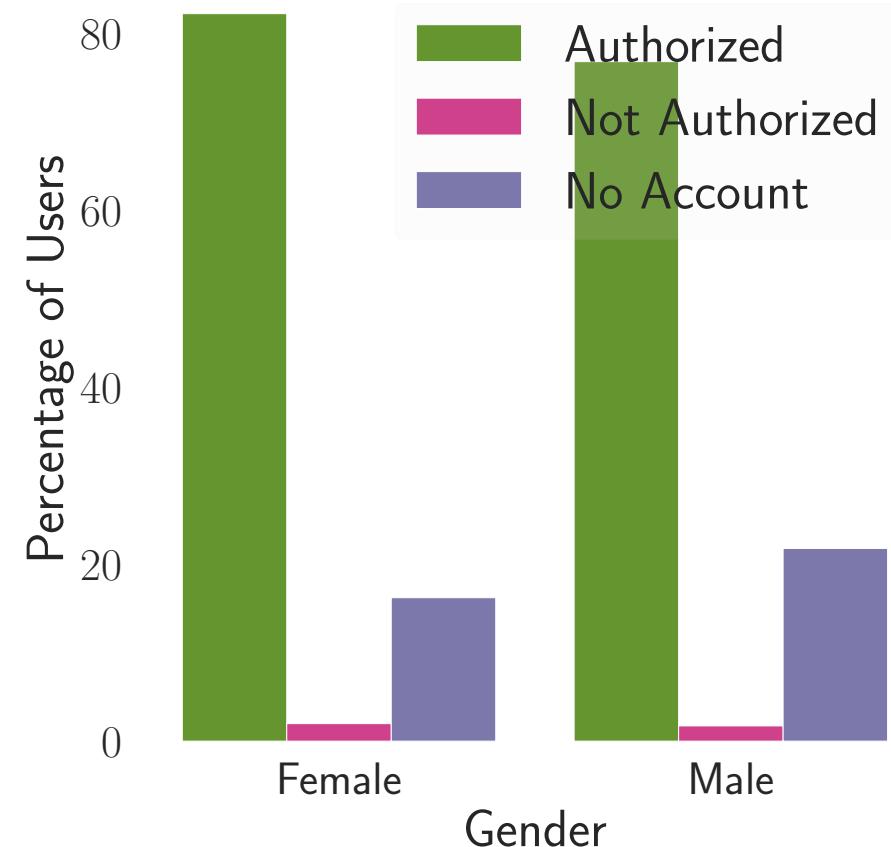
Out of 757 participants (Facebook data):

- 587 participants consented and **authorized**
- 67 consented, **did not** authorize
- 103 **did not have** Facebook Account

We did **not** collect **photos, media** and **private messages**

Who authorized Facebook Data?

- ▶ Female participants more likely to authorize than Males
- ▶ Male participants less likely to have a Facebook Account
- ▶ No significant difference in authorization behavior across age, income, personality trait

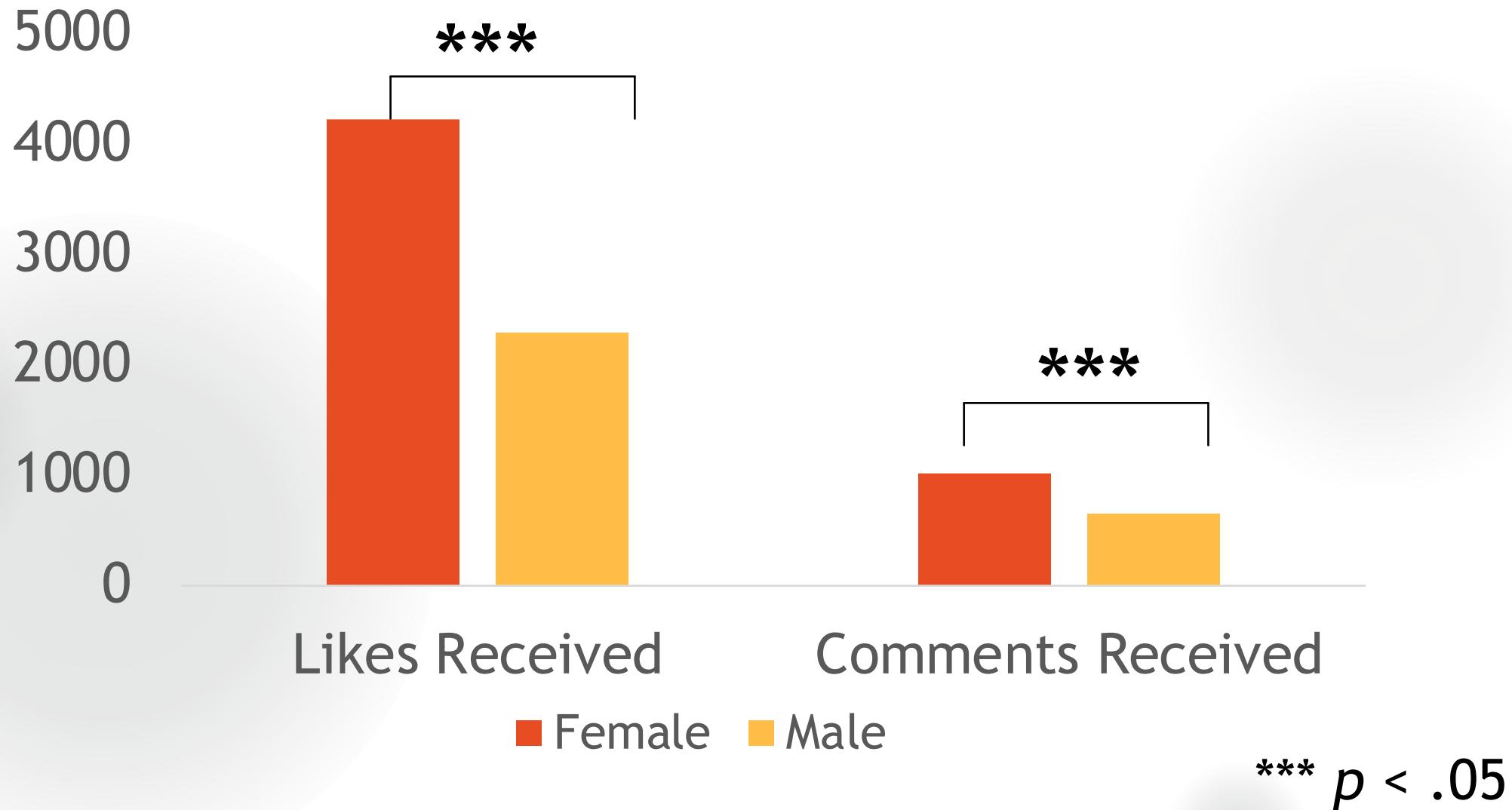


Descriptive Statistics

- ▶ 237,725 Timeline updates (Median: 195 per participant)
- ▶ 1,672,482 Likes received (Median: 1,1,51 per participant)
- ▶ 452,003 Comments received (Median: 331 per participant)
- ▶ 1,917 days of data on an average per participant (October 2005 - August 2018)



Associating with Participant Attributes: Demographic

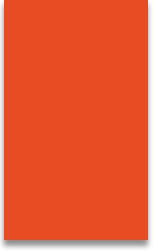


Associating with Participant Attributes: Personality

- ▶ Higher agreeableness / extraversion / openness is associated with **greater likes and comments received**
- ▶ Higher conscientiousness is associated with **shorter posts**
- ▶ Higher neuroticism / openness is associated with **longer posts**

Associating with Participant Attributes: Wellbeing

- ▶ Poorer sleep quality is associated with longer posts, more likes and more comments
- ▶ Higher negative affect is associated with lesser likes received



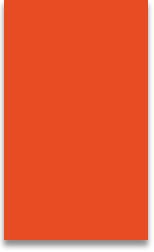
Takeaways, Lessons, and Guidelines



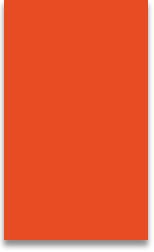
An **infrastructure** to unobtrusively
collect social media data at scale



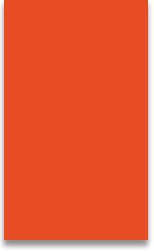
Who agrees **to share** their social media data for research?



Observation: Differences in data with
personality traits and **wellbeing attributes**



Recommendation: Control for **gender** in terms of social media **data quantity**



We share **de-identified** (and specially
consented) sample of our dataset for
research purposes

Koustuv Saha, Ayse E. Bayraktaroglu, Andrew T. Campbell, Nitesh V. Chawla, Munmun De Choudhury, Sidney K. D'Mello, Anind K. Dey, Ge Gao, Julie M. Gregg, Krithika Jagannath, Gloria Mark, Gonzalo J. Martinez, Stephen M. Mattingly, Edward Moskal, Anusha Sirigiri, Aaron Striegel, and Dong Whi Yoo. 2019. Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing. In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19). ACM, New York, NY, USA, Paper CS12, 8 pages. DOI: <https://doi.org/10.1145/3290607.3299065>

Social Media as a Passive Sensor in Longitudinal Studies of Human Behavior and Wellbeing

This research is supported in part by the Office of the Director of National Intelligence (ODNI),
Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2017-17042800007.



Thank You

@kous2v | koustuv.saha@gatech.edu | koustuv.com