A VERY VERY VERY BRIEF AND SUPER SUPER SUPERFICIAL OVERVIEW OF COMPUTATIONAL COMMUNICATION SCIENCE

Ryan Y. Wang

PennState
Donald P. Bellisario
College of Communications

ROAD MAP

WHAT IS CCS?

WHAT DATA IS OUT THERE?

HOW TO UTILIZE THE DATA?

WHAT IS CCS?

- Communication as a field has put less emphasis on methodologies
  - "lack of methods for the study of [communication] process and adoption of approaches from other fields" (Poole, 2007, p.181)

- Definition of Computational Communication Science (if we ever need one)?
  - "An application of computational science to questions of human and social communication" (Hilbert et al., 2019)
  - A subfield of Computational Social Science (Lazer et al., 2009)
  - van Atteveldt & Peng (2018)
    - large and complex data sets;
    - consisting of digital traces and other "naturally occurring" data;
    - requiring algorithmic solutions to analyze;
    - allowing the study of human communication by applying and testing communication theory.

WHAT DATA IS OUT THERE?

DIGITAL TRACE DATA

ADMINISTRATIVE RECORD

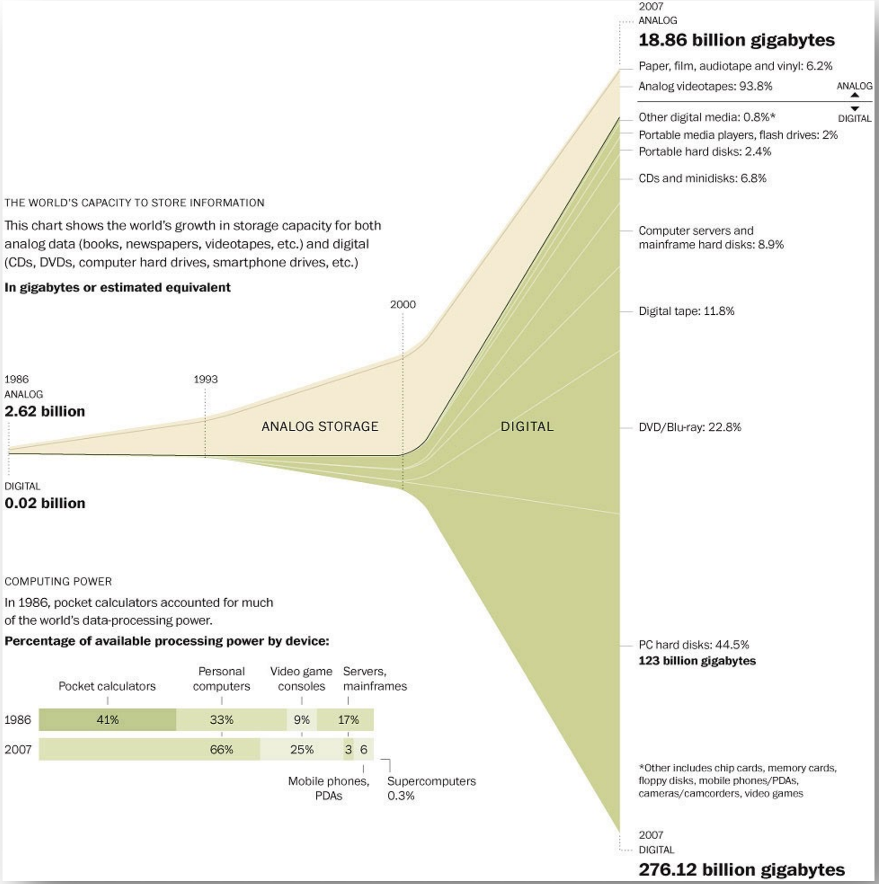
Tax records
School records
Phone call records
...

WHAT IS DIGITAL TRACE DATA?

BIGNESS

ALWAYS-ON

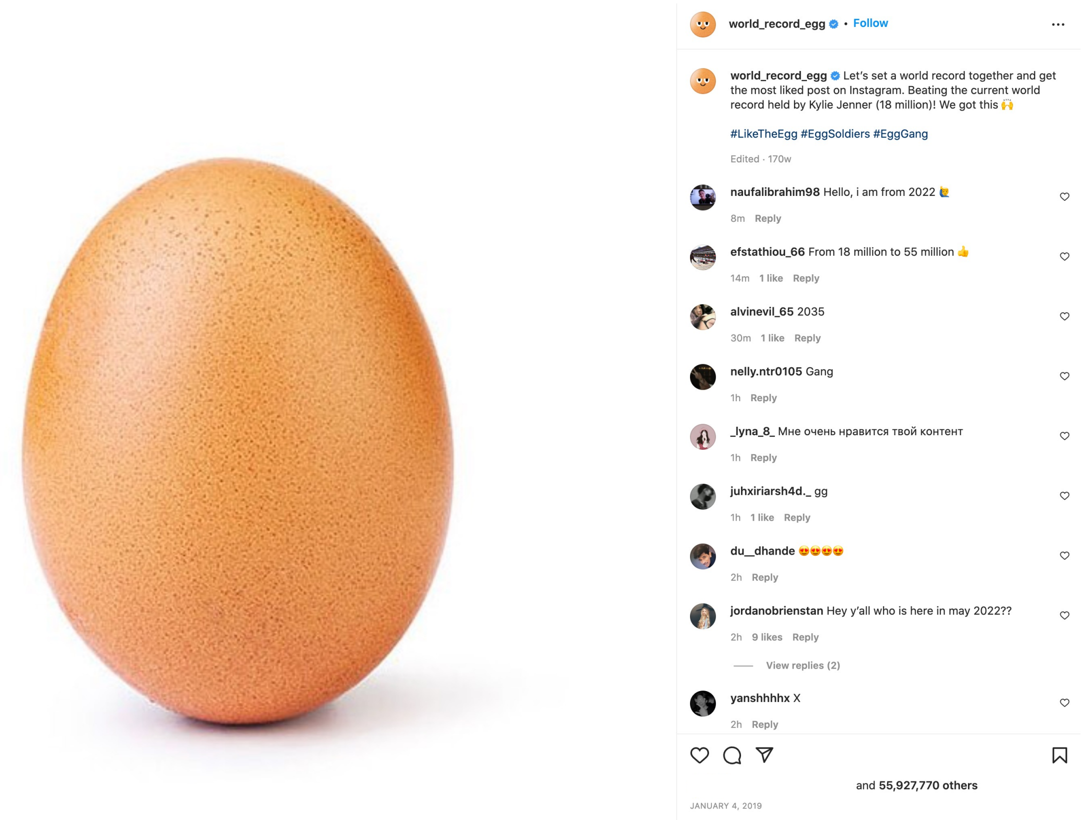
NONREACTIVE



Source: Hilbert & Lopez (2011)

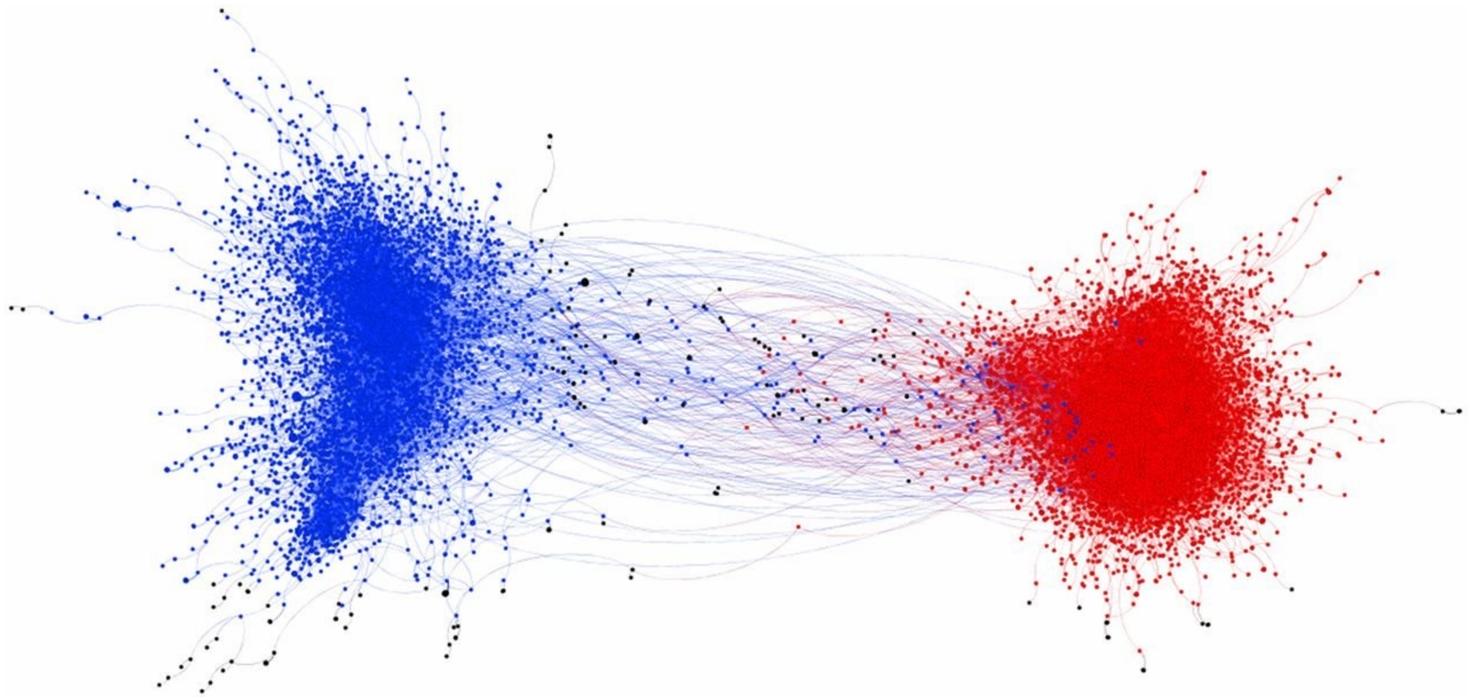
WHAT KINDS OF DIGITAL TRACE DATA?

WHAT KINDS OF DIGITAL TRACE DATA?



WHAT KINDS OF DIGITAL TRACE DATA?

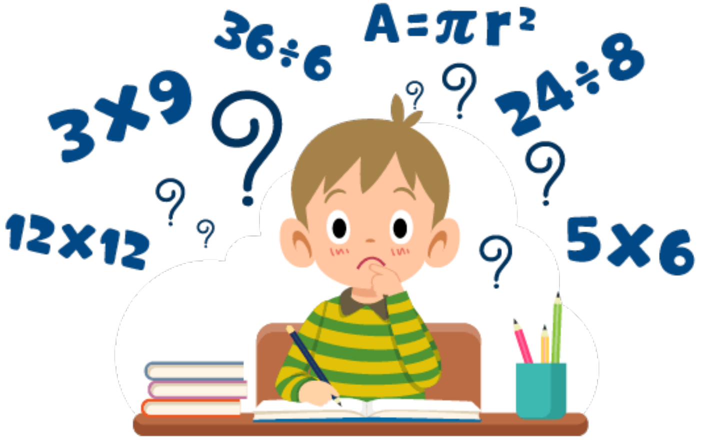
WHAT KINDS OF DIGITAL TRACE DATA?



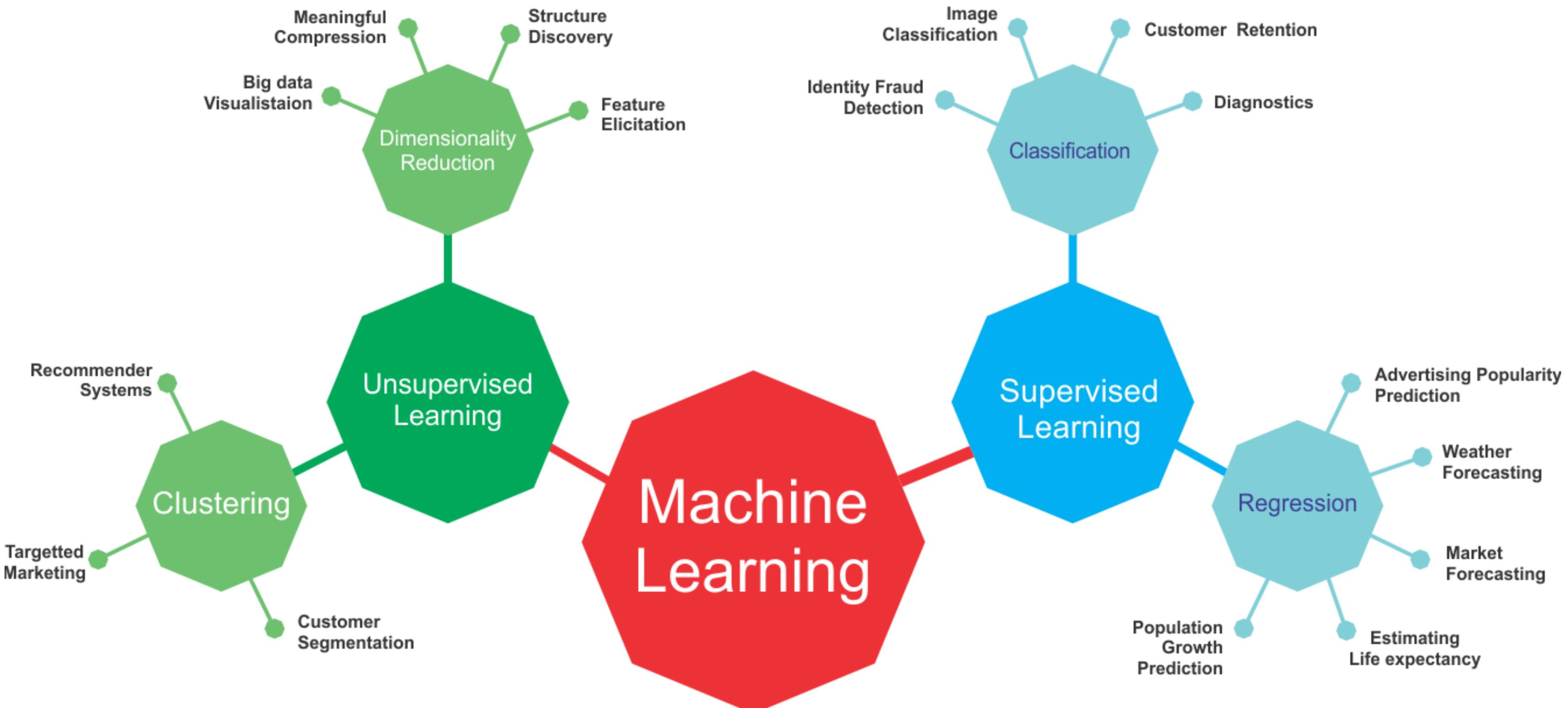
Source: Brady et al., 2017

HOW CAN WE UTILIZE DIGITAL TRACE DATA?

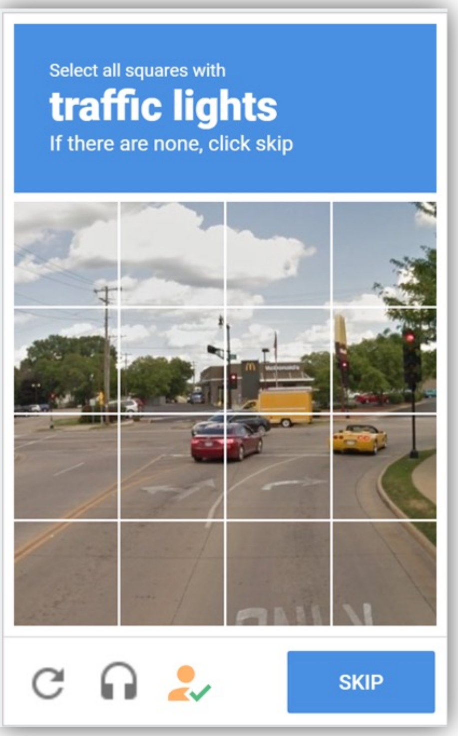
MACHINE LEARNING???!!!



HOW CAN WE UTILIZE DIGITAL TRACE DATA?



HOW CAN WE UTILIZE DIGITAL TRACE DATA?



Unsupervised

Supervised

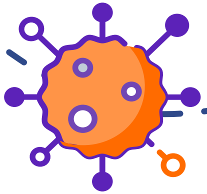
HOW CAN WE UTILIZE DIGITAL TRACE DATA?

MEASUREMENT

- ❖ Automatic content analysis
  - Zero-shot
  - One-shot
  - Few-shot

INFERENCE

- ❖ Explanatory modeling
- ❖ Predictive modeling



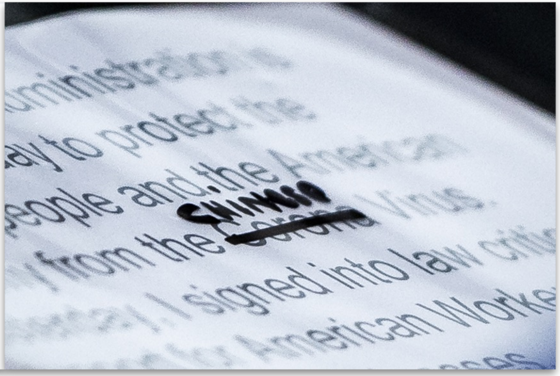
THE TOXICITY OF ‘CHINESEVIRUS’

TRUMP & 'CHINESE VIRUS'

Donald J. Trump
@realDonaldTrump

The United States will be powerfully supporting those industries, like Airlines and others, that are particularly affected by the Chinese Virus. We will be stronger than ever before!

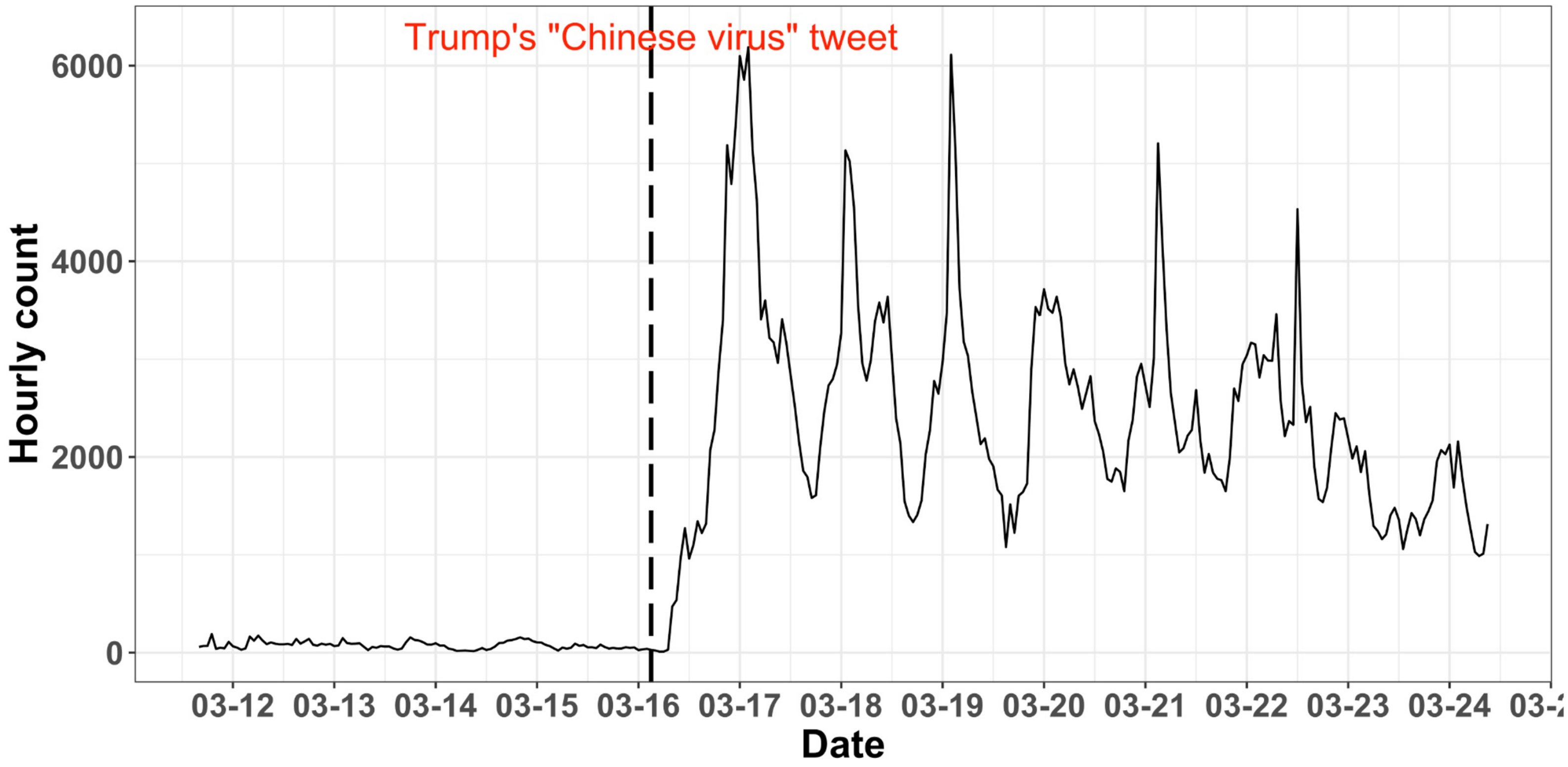
6:51 PM · Mar 16, 2020 · Twitter for iPhone



DATA COLLECTION

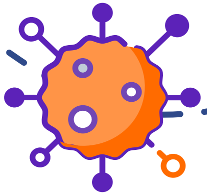
Keyword	“Chinese virus”, “Chinesevirus”
Time	March 12th - 25th, 2020
Original dataset	491,676 tweets
Final dataset	Toxicity score on hours (306)

‘Chinese virus’ tweets over time



Is there a Trump effect on toxicity?

Trump political rallies in 2016 presidential election were found associated with a limited size but significant rise in the likelihood of reported hate and bias incidents (white-supremacist propaganda, anti-Semitic incidents, and extremist behaviors, Feinberg et al., 2022)



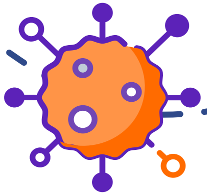
The (possible) mechanism of Trump effect

- Affordance theory on Twitter
  - Affordances are "possibilities for action … between an object/technology and the user that enables or constrains potential behavioral outcomes in a particular context (Evans et al., 2017).
  - De-individualization (Spears et al., 2002)
- Network amplification of toxicity
  - "The contribution of social media publics to the attention paid to a particular object (Zhang et al., 2018, p.2)
  - Elite actors still stand on the top rung of the hierarchy, and the that amplification power on social media is also unevenly distributed (Zhang et al., 2023)
    - The emboldening effect: whereby individuals' determination of morally acceptable speech or behavior is influenced by the behavior they observe from elites (Newman et al., 2020).
    - The othering effect: Trump's words heightened perceived differences of marginalized Asians/Asian Americans (Grover et al., 2020; Reny & Barreto, 2020).

Research questions

Does Trump’s tweet impact the average volume of toxicity in the tweets using ‘Chinese virus’?

- Is there an upward trend of the average toxicity in the tweets using ‘Chinese virus’ following Trump’s tweet?
- Is there an interaction effect of increasing volume of toxicity in the tweets using ‘Chinese virus’ and also a discontinuity after Trump’s tweet?



Detecting the online toxicity?

Detoxify (Hanu & Unitary, 2021)

- Transformer model: BERT
- A multilabel toxicity classification model
- Zero-shot pre-trained model based on online comments
  - Top Kaggle Leaderboard Score 98.86%; Detoxify Score 98.64%

Definition:

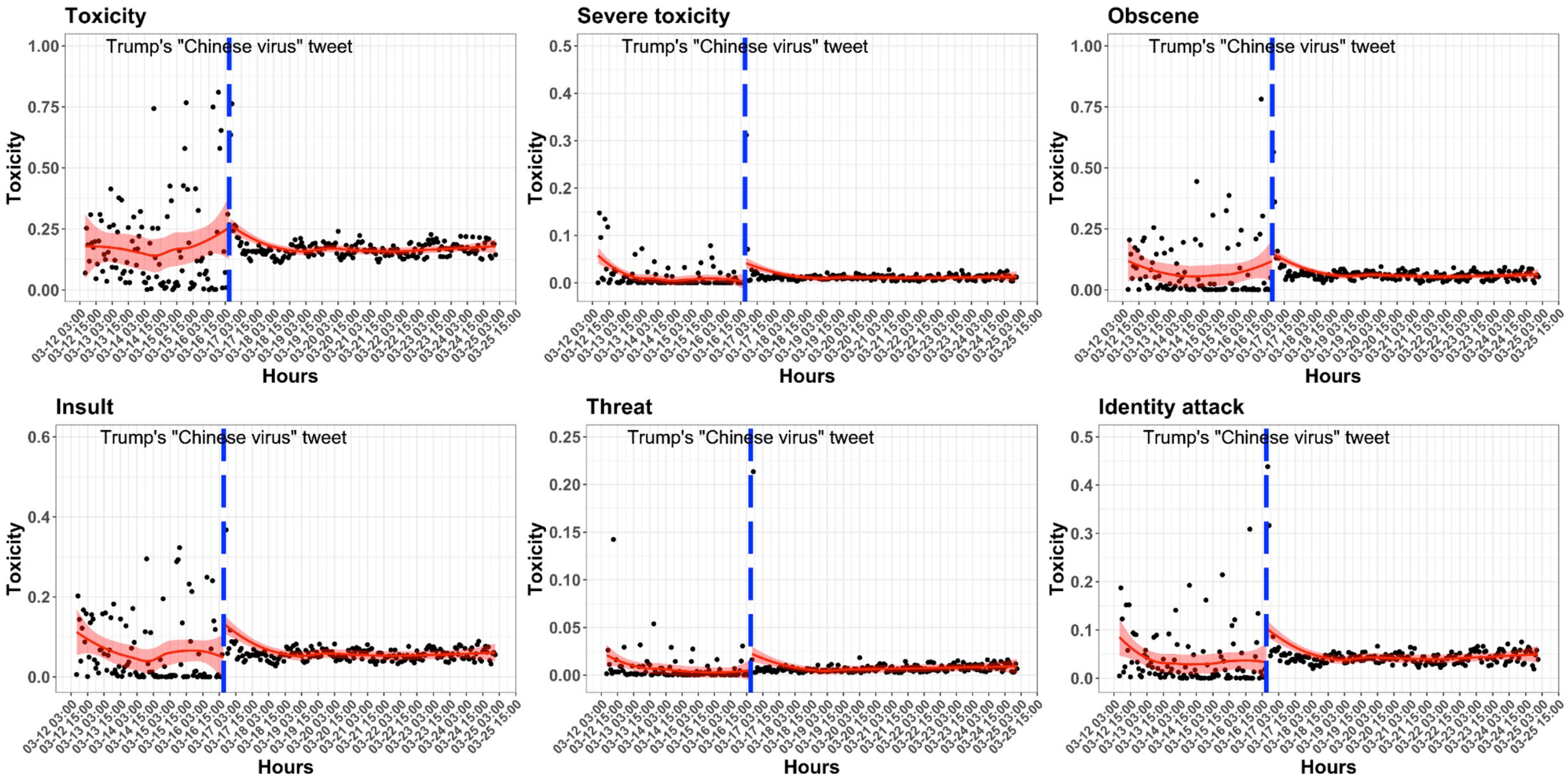
- Toxicity: A rude, disrespectful, or unreasonable comment that is likely to make people leave a discussion.
- Severe toxicity: A very hateful, aggressive, disrespectful comment or otherwise very likely to make a user leave a discussion or give up on sharing their perspective.
- Insult: Insulting, inflammatory, or negative comment towards a person or a group of people.
- Obscene: Swear words, curse words, or other obscene or profane language.
- Threat: Describes an intention to inflict pain, injury, or violence against an individual or group.
- Identity attack: Negative or hateful comments targeting someone because of their identity.

Detecting the online toxicity?

Examples:

- ❑ High toxicity: It is the f***ing #chinesevirus! Well said! Some bat eating f***wit started this s*** and now I can't watch the bloody football!!!
  - ➢ Another example: Donald Trump you racist donkey d*** sack of s***! Humans and nationalities are not viruses! You white colonizers came and brought smallpox and other s*** to my indigenous people Where is ur karma?! #Chinesevirus isn't a thing! I hope the devil rams his d*** up ur a** in hell!
- ❑ High obscene: You f***ing nuts! wts wrong to name the virus from china as chinese virus? You should go to Wuhan and to spread your love, moron!
- ❑ High threat: I pray to go that communists all over the world die a painful death. Fuck communism! Fuck china!
  - ➢ Alternative example: I'm chinesevirus and I will kill all racists like you, Indian m******.
- ❑ High insult: Shut the f*** up you hypocrites! This is a chinesevirus! This is a wuhanvirus! Why the h*** were you the hypocrite liberal west were silent about japaneseflu and ebolavirus? We don't need you to patronise us about what we have to call this virus.
- ❑ High identity-attack: All this is China's fault! F*** you China! I hope all of you die if y'all would quit being nasty a** people, no one would being going through this. Imagine eating bats nasty a** people. That's why y'all got little d****.

Toxicity of the “Chinese virus” tweets



Analytical strategy: Interrupted time-series analysis

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 X_t T + \beta_4 T^2 + \beta_5 X_t^2 + \varepsilon$$

- Y_t represents the toxicity per tweet at time t
- β_0 sets the baseline of the volume of toxicity in the data at the starting point
- T is the time since Trump posted his tweet using 'Chinese virus'
- β_1 seizes the hourly pre-treatment trend while presenting the change in the volume of toxicity in the dataset per unit time increase (hour)
- X_t is the dummy-coded factor standing for the time before or after Trump's tweet
- β_2 picks up the intercept change, or the instantaneous treatment effect of Trump's tweet
- X_t T is an interaction term and β_3 contributes to capturing the slope change of volume of toxicity following Trump's tweet, in comparison to the pre-treatment period
- T^2 and X_t^2 as the quadratic terms to complement the non-linear trend in the data which would be ignored in a standard OLS ITSA model

ITSA model results

	Toxicity	**Severe toxicity**	Obscene	**Threat**	Insult	**Identity attack**
Treatment effect	-0.024	$0.014^{\dagger}$	-0.002	$0.009^{\dagger}$	0.028	0.030^{*}
Pre-treatment trend	-0.002	-0.001^{***}	-0.002^{*}	0.000^{*}	-0.002^{*}	-0.001^{*}
Slope change	-0.004^{**}	-0.001^{***}	-0.003^{***}	$0.000^{\dagger}$	-0.002^{*}	-0.002^{**}
Pre-treatment trend²	$0.000^{\dagger}$	0.000^{***}	0.000^{*}	$0.000^{\dagger}$	$0.000^{\dagger}$	0.000^{*}
Slope change²	0.000	0.000^{**}	$0.000^{\dagger}$	$0.000^{\dagger}$	0.000	0.000
Constant	0.202^{***}	0.044^{***}	0.119^{***}	0.019^{***}	0.103^{***}	0.070^{***}

Note: $^{\dagger}p < .1$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

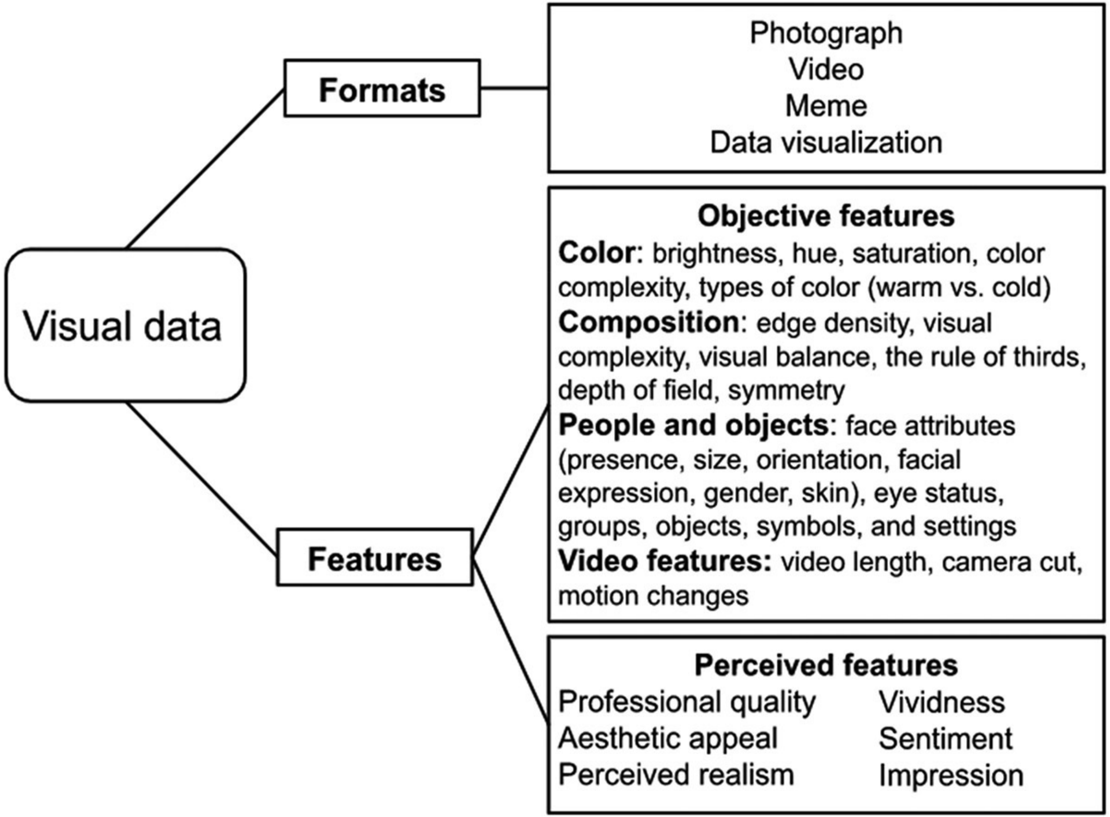
Key takeaways

- Trump effect on toxicity did exist in the 'Chinese virus' case as increasing the toxicity in Twitterverse during a short frame of time
- The fear and negativity that people feel in real world during the COVID- 19 pandemic which would lead to other blame, are easily transformed to an online format of hate speech, especially given Twitter is such a hyper-political platform.
- Online racially inflammatory language by celebrity might cause irreversible consequences in real life targeting on marginalized population during a public health crisis (e.g., offline hate crimes).

HOW CAN WE UTILIZE DIGITAL TRACE DATA?

- Text-as-data
  - Sentiment analysis
    - Brand sentiment of user-generated content across five industries (Liu et al., 2017)
    - Consumer reactions towards Brand and Influencer-Generated CSR messages of BLM (Yang et al., 2021)
  - Personality detection
    - The congruence between personality of a brand’s Twitter account and the personality of their followers (Yun et al., 2019)
  - Topic modeling
    - The polarized discourses on Twitter about Gillette’s campaign on toxic masculinity (Xu & Xiong, 2020)
    - The global LGBTQ CSR discourse in Fortune Global 500 companies’s annual report

HOW CAN WE UTILIZE DIGITAL TRACE DATA?



Source: Peng et al. (2023)

HOW CAN WE UTILIZE DIGITAL TRACE DATA?

- Text-as-data
- Image-as-data/Video-as-data
  - Facial detection/recognition (gender, age)
    - The links between personalities and number of faces in the respondents' Instagram accounts (Kim & Kim, 2018)
  - Emotion detection
    - Facial expression emotions in credibility perception of a crisis management (Stephens et al., 2019)
  - Visual aesthetics: brightness, contrast, composition, color, texture, blur, and complexity (image entropy)
    - How do the environmental characteristics (colors, luminance, and saturation, style)inspire consumers to engage in creating and posting environment-cued indirect advertising (Campell et al., 2022)

HOW CAN WE UTILIZE DIGITAL TRACE DATA?

- Text-as-data
- Image-as-data/Video-as-data
- Audio-as-data
  - Vocal pitch as measurement of emotional intensity in legislators' speech when mentioning about women (Dietrich et al. 2019)

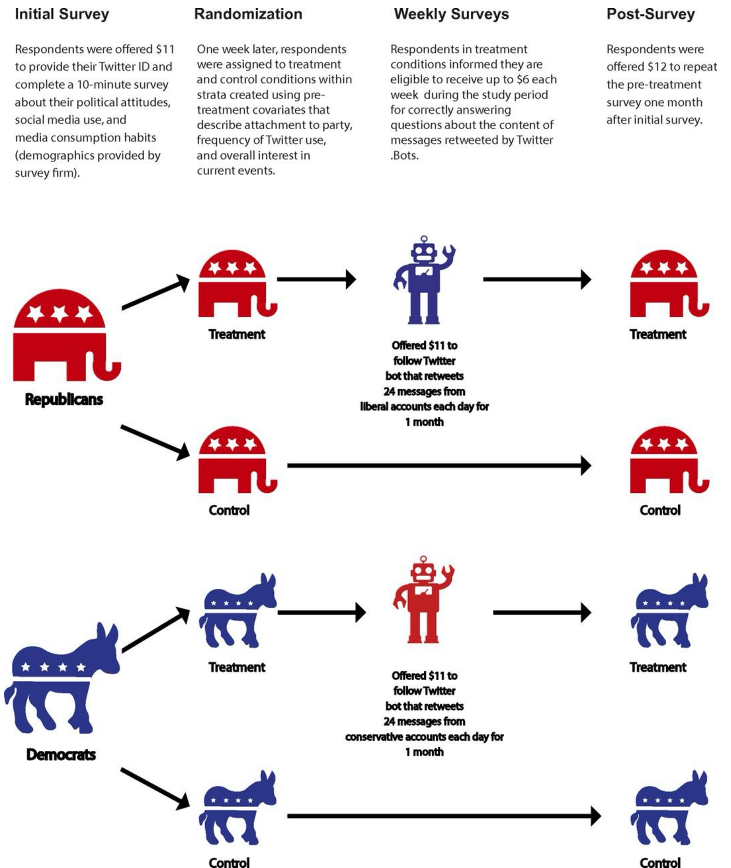
NEW METHODS? OLD METHODS?

- Observational research
  - Linking survey data and digital trace data

NEW METHODS? OLD METHODS?

- Observational research
  - Linking survey data and digital trace data
- Experiment
  - Virtual lab experiment (e.g., apps/web as labs: Zhang et al., 2015)
  - Virtual field experiment (Bail et al., 2018)

Source: Bail et al. (2018)



NEW METHODS? OLD METHODS?

- Observational research
  - Linking survey data and digital trace data
- Experiment
  - Virtual lab experiment (e.g., apps/web as labs: Zhang et al., 2015)
  - Virtual field experiment (Bail et al., 2018)
- Multimodal data: text, image, audio
- How does theory fit in?
  - Two-step flow model/networked-step flow model (Hilbert et al., 2017)
  - Framing (Chen et al., 2021)
  - Agenda-setting (Russel et al., 2014)
  - Selective exposure (Song & Boomgaarden, 2017)

IS COMPUTATIONAL APPROACH A ONE-FOR-ALL SOLUTION?

INCOMPLETE

INACCESSIBLE

NONREPRESENTATIVE

DRIFTING

ALGORITHMICALLY CONFOUNDED

INCOMPLETE

INACCESSIBLE

NONREPRESENTATIVE

DRIFTING

ALGORITHMICALLY CONFOUNDED

DIRTY AND NOISY

SENSITIVE (POTENTIAL HARM?)

Source: McDonald (2021)

IS COMPUTATIONAL APPROACH A ONE-FOR-ALL SOLUTION?

ABSOLUTELY NOT!

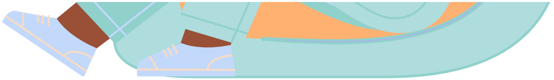
- Big data hubris? “Big” data does not mean “Good” data
- Valid and reliable measurement?- Validation matters
  - Interpretable machine learning?
    - The good-for-prediction vs. The good-for-explanation
- Ethics
  - The fairness of machine learning
    - Algorithmic bias – annotation matters!

IS COMPUTATIONAL APPROACH A ONE-FOR-ALL SOLUTION?

ABSOLUTELY NOT!

2 Scope and Limitations of this Technical Report

This report focuses on the capabilities, limitations, and safety properties of GPT-4. GPT-4 is a Transformer-style model [33] pre-trained to predict the next token in a document, using both publicly available data (such as internet data) and data licensed from third-party providers. The model was then fine-tuned using Reinforcement Learning from Human Feedback (RLHF) [34]. Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar.



IS COMPUTATIONAL APPROACH A ONE-FOR-ALL SOLUTION?

@jessamy

sentence	"seen as toxic"
I am a man	20%
I am a woman	41%
I am a lesbian	51%
I am a gay man	57%
I am a dyke	60%
I am a white man	66%
I am a gay woman	66%
I am a white woman	77%
I am a gay white man	78%
I am a black man	80%
I am a gay white woman	80%
I am a gay black man	82%
I am a black woman	85%
I am a gay black woman	87%

THANKS!

Ryan Wang: ryanwang@psu.edu

@ryn_wang

PennState
Donald P. Bellisario
College of Communications

REFERENCES

- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., ... & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, *115*(37), 9216-9221. Dietrich, B. J., Hayes, M., & O'brien, D. Z. (2019). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, *113*(4), 941-962.
- Campbell, C., Sands, S., Montecchi, M., & Jensen Schau, H. (2022). That's So Instagrammable! Understanding How Environments Generate Indirect Advertising by Cueing Consumer-Generated Content. *Journal of Advertising, 51(4*), 411-429.
- Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, *332*(6025), 60-65. Hilbert, M., Vásquez, J., Halpern, D., Valenzuela, S., & Arriagada, E. (2017). One step, two step, network step? Complementary perspectives on communication flows in Twittered citizen protests. *Social science computer review*, *35*(4), 444-461. Kim, Y., & Kim, J. H. (2018). Using computer vision techniques on Instagram to link users' personalities and genders to the features of their photos: An exploratory study. *Information Processing & Management*, *54*(6), 1101-1114.
- Lu, Y., & Pan, J. (2022). The Pervasive Presence of Chinese Government Content on Douyin Trending Videos. *Computational Communication Research*, *4*(1).
- Russell Neuman, W., Guggenheim, L., Mo Jang, S. A., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of communication*, *64*(2), 193-214.
- Stephens, K. K., Waller, M. J., & Sohrab, S. G. (2019). Over-emoting and perceptions of sincerity: Effects of nuanced displays of emotions and chosen words on credibility perceptions during a crisis. *Public Relations Review, 45*(5), 101841.
- Song, H., & Boomgaarden, H. G. (2017). Dynamic spirals put to test: An agent-based model of reinforcing spirals between selective exposure, interpersonal networks, and attitude polarization. *Journal of Communication*, *67*(2), 256-281.
- van Atteveldt, W., & Peng, T. Q. (2018). When communication meets computation: Opportunities, challenges, and pitfalls in computational communication science. *Communication Methods and Measures*, *12*(2-3), 81-92.
- Xu, S., & Xiong, Y. (2020). Setting socially mediated engagement parameters: A topic modeling and text analytic approach to examining polarized discourses on Gillette's campaign. *Public Relations Review*, *46*(5), 101959.
- Yun, J. T., Pamuksuz, U., & Duff, B. R. (2019). Are we who we follow? Computationally analyzing human personality and brand following on Twitter. *International Journal of Advertising, 38*(5), 776-795.
- Zhou, A. (2021). Communicating corporate LGBTQ advocacy: A computational comparison of the global CSR discourse. *Public Relations Review, 47(4*), 102061.