Twitter: @BostonUResearch | #researchontap

RESEARCH ON TAP

Toward Responsible AI: Privacy, Fairness, and Accountability

Wednesday, October 25, 2023

bu.edu/research/events

BOSTON UNIVERSITY

Boston University Office of Research

Agenda

- Welcome Remarks
- Presentations
  - Stacey Dogan & Woody Hartzog
  - Christopher Robertson
  - Andy Sellars
  - Mark Crovella
  - Tesary Lin
  - Chris Chao Su
  - Adam Smith
  - Marshall Van Alstyne
- Closing Remarks

Hariri Institute

CENTERS AND INITIATIVES

MOC Alliance

Software & Application Innovation Lab (SAIL)

Center for Computational Science

Red Hat Collaboratory

AI and Education Initiative

Center for Reliable Information Systems and Cyber Security

AI Research Initiative (AIR)

Boston Women's Workforce Council

Center for Information and Systems Engineering (CISE)

Digital Health Initiative

Funding Opportunity: Focused Research Program

To support medium-large interdisciplinary teams, combining senior PI leadership and junior PI talents, helping them coalesce around an exciting emerging area, enabling societal impact, and preparing them for future large sponsored programs.

Proposing a Focused Research Program is a two-stage process. Pre-proposals are due by 1/29/24. FRP Proposals are due 3/8/24.

Faculty interested in submitting a proposal are strongly encouraged to discuss their ideas with the Hariri Institute's Director, **Yannis Paschalidis, yannisp@bu.edu**.

Hariri Institute for Computing Faculty Affiliate Program

The Hariri Institute for Computing community consists of **380+ Faculty Affiliates** from 66 departments, across 13 Boston University Schools and Colleges. The **program is open to all BU faculty members pursuing research projects, or leading teaching or training initiatives in computing and/or computational science & engineering.**

Benefits Including:

- Opportunities to network and collaborate across and beyond Boston University with academic, industry, and government researchers.
- Promote noteworthy achievements through Institute portals and social media.
- Pre/post-award grant administration.
- Access to Institute hybrid meeting and event spaces.

How to apply?

- Fill out an application: https://www.bu.edu/hic/2019-hariri-affiliate-program-application/

Connect with us!

Hariri Institute for Computing, Boston University

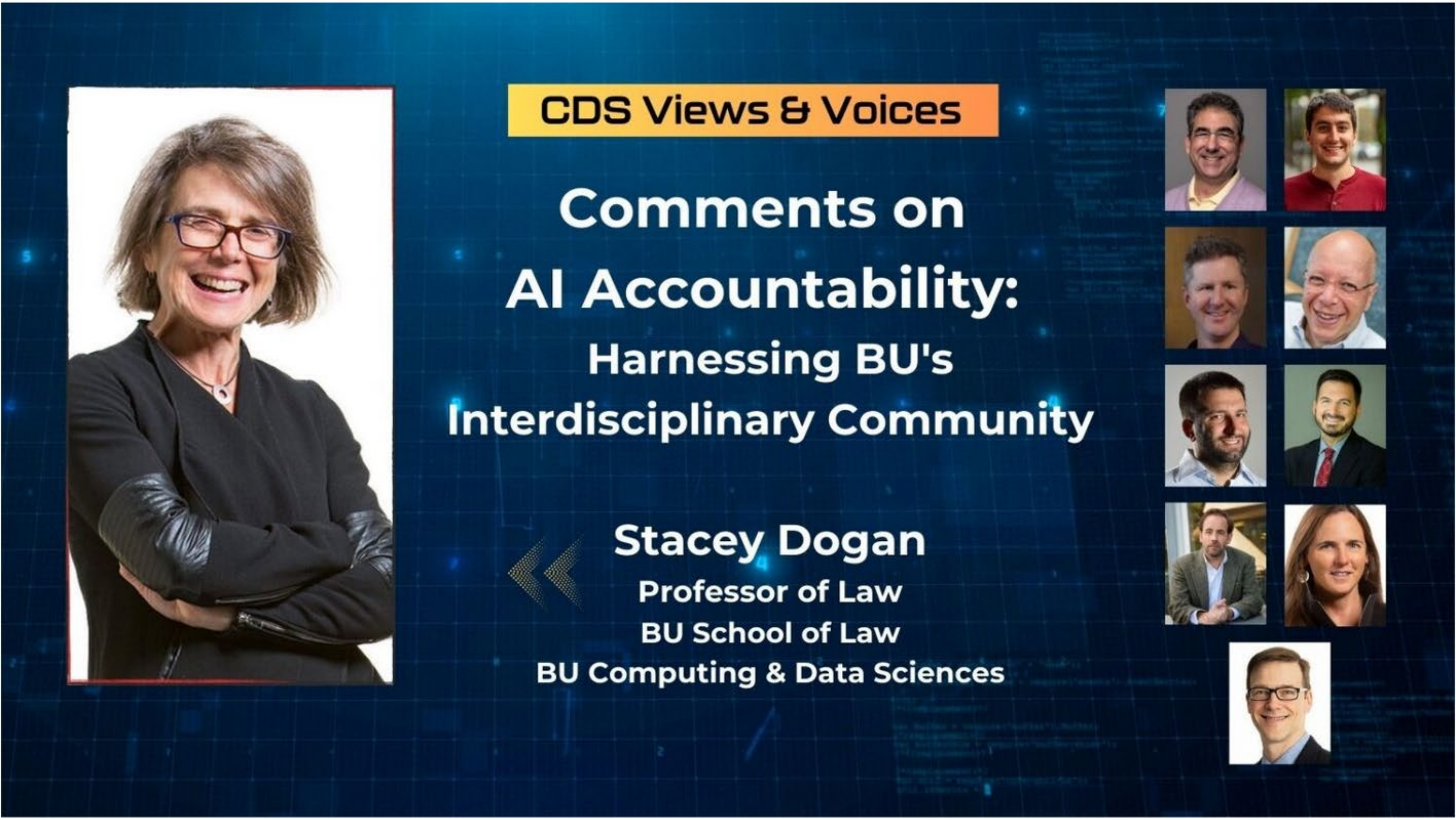
@BU_Computing

@BU_Computing

Hariri Institute for Computing, Boston University

What's Happening With Responsible AI at BU?

Stacey Dogan
Woodrow Hartzog



CDS Views & Voices

Comments on AI Accountability:

Harnessing BU's Interdisciplinary Community

Stacey Dogan

Professor of Law

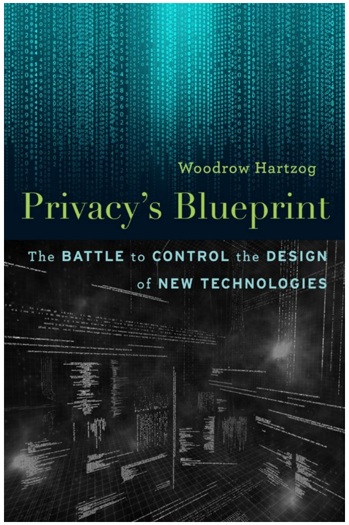
BU School of Law

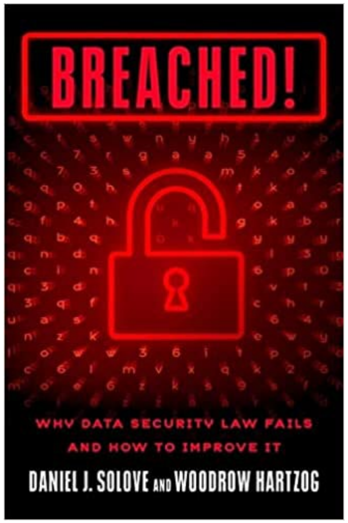
BU Computing & Data Sciences

Woodrow Hartzog

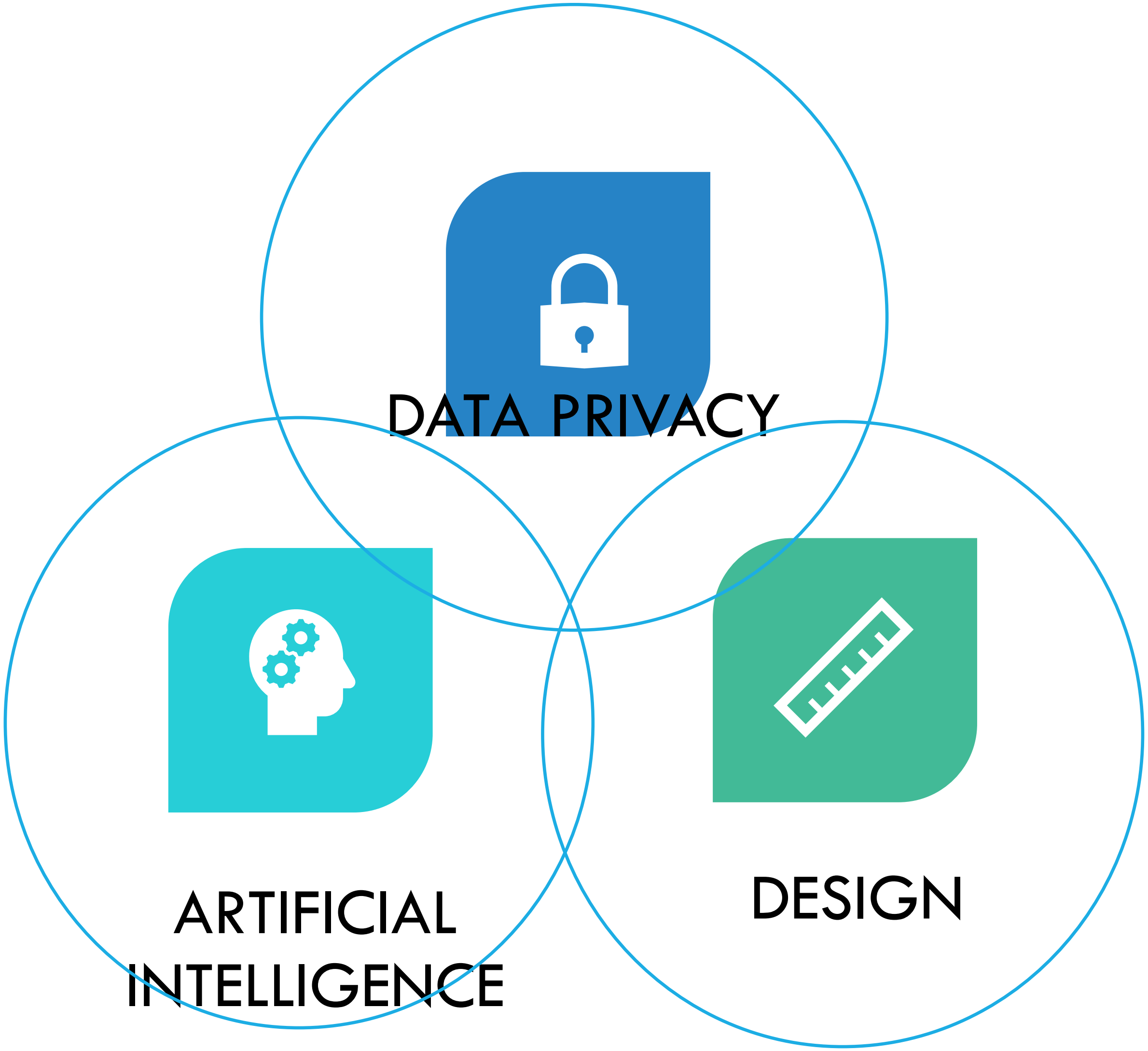
Professor of Law **Boston University School of Law**

- Law
- Robotics and artificial intelligence
- Privacy





BREACHED!
WHY DATA SECURITY LAW FAILS AND HOW TO IMPROVE IT
DANIEL J. SOLOVE AND WOODROW HARTZOG

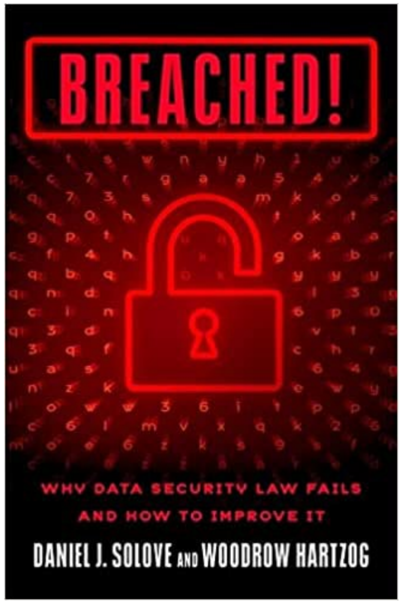


DATA PRIVACY

ARTIFICIAL INTELLIGENCE

DESIGN

Privacy (Trust, Obscurity & Design)

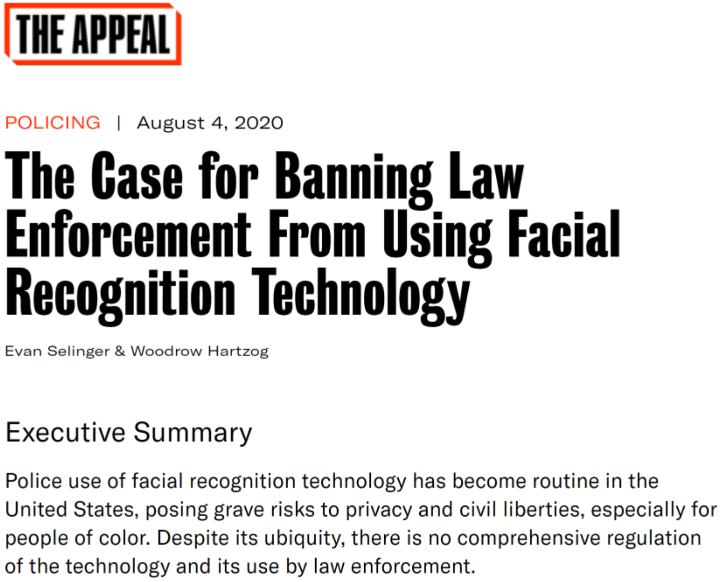


BREACHED!

WHY DATA SECURITY LAW FAILS AND HOW TO IMPROVE IT

DANIEL J. SOLOVE AND WOODROW HARTZOG

Artificial Intelligence



THE APPEAL

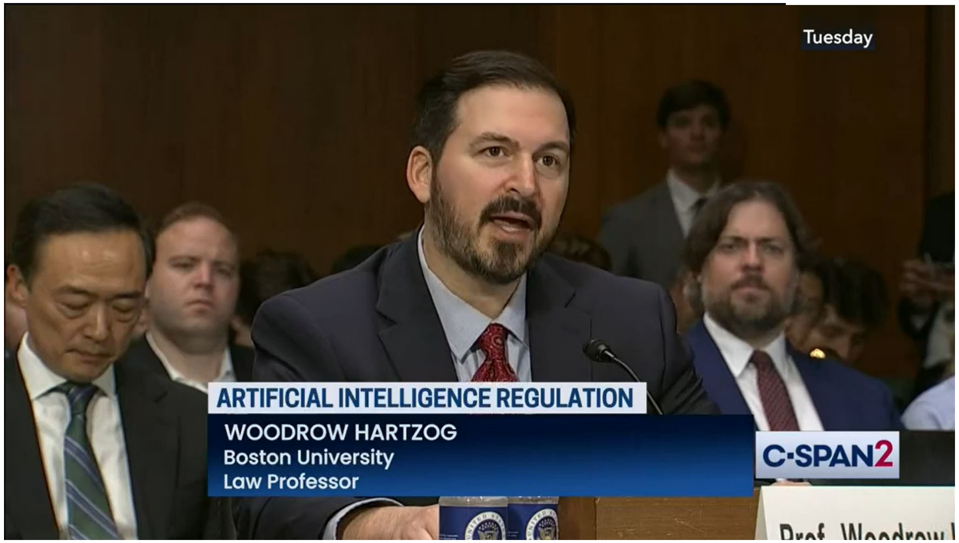
POLICING | August 4, 2020

The Case for Banning Law Enforcement From Using Facial Recognition Technology

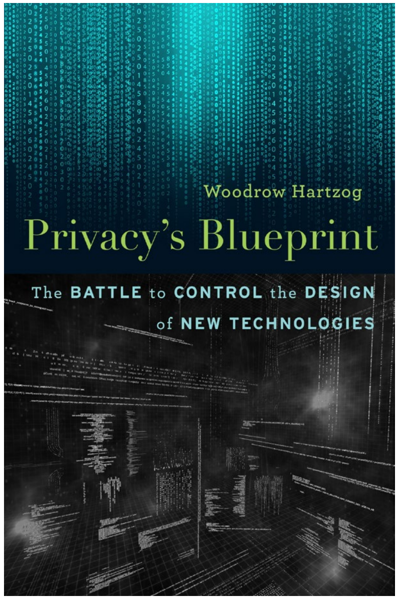
Evan Selinger & Woodrow Hartzog

Executive Summary

Police use of facial recognition technology has become routine in the United States, posing grave risks to privacy and civil liberties, especially for people of color. Despite its ubiquity, there is no comprehensive regulation of the technology and its use by law enforcement.



Design

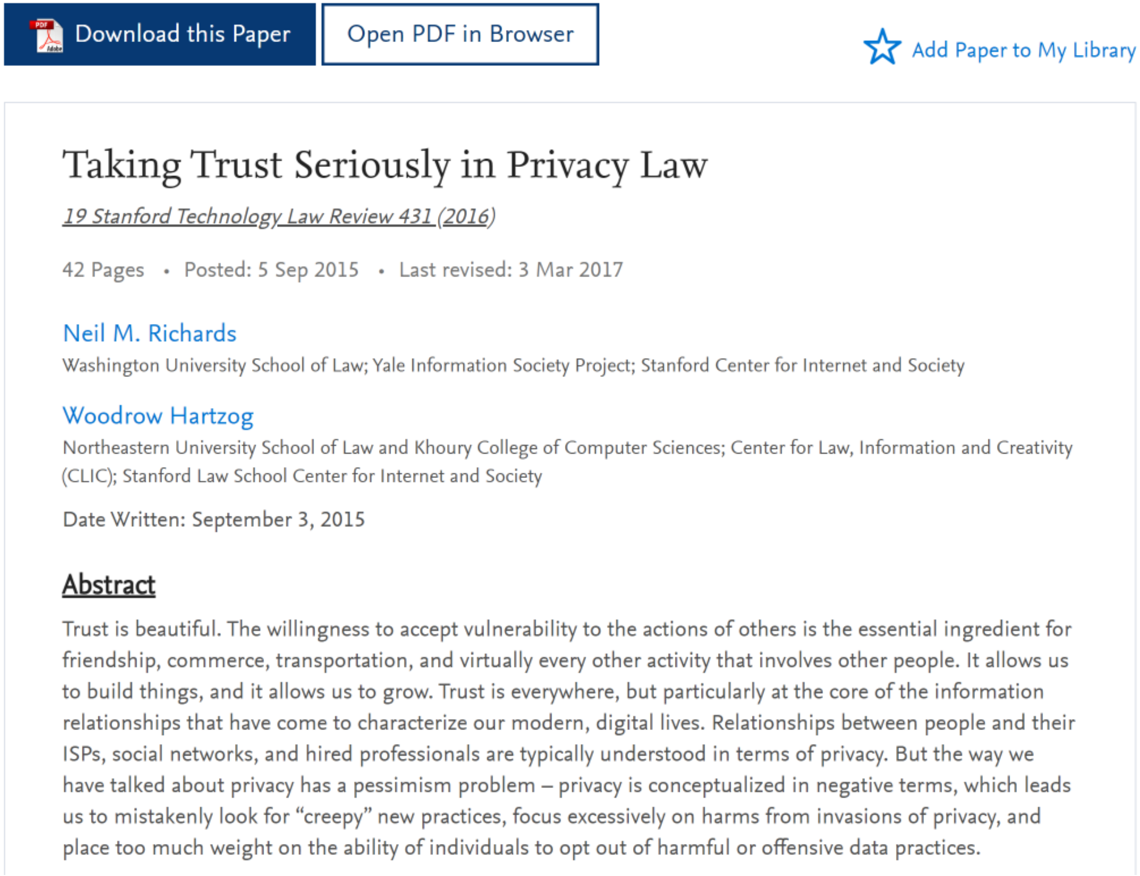

Woodrow Hartzog

Privacy's Blueprint

The BATTLE to CONTROL the DESIGN of NEW TECHNOLOGIES

FEDERAL TRADE COMMISSION ★ UNITED STATES OF AMERICA ★ MCMXV ★

TRUST



Taking Trust Seriously in Privacy Law

19 Stanford Technology Law Review 431 (2016)

42 Pages • Posted: 5 Sep 2015 • Last revised: 3 Mar 2017

Neil M. Richards
Washington University School of Law; Yale Information Society Project; Stanford Center for Internet and Society

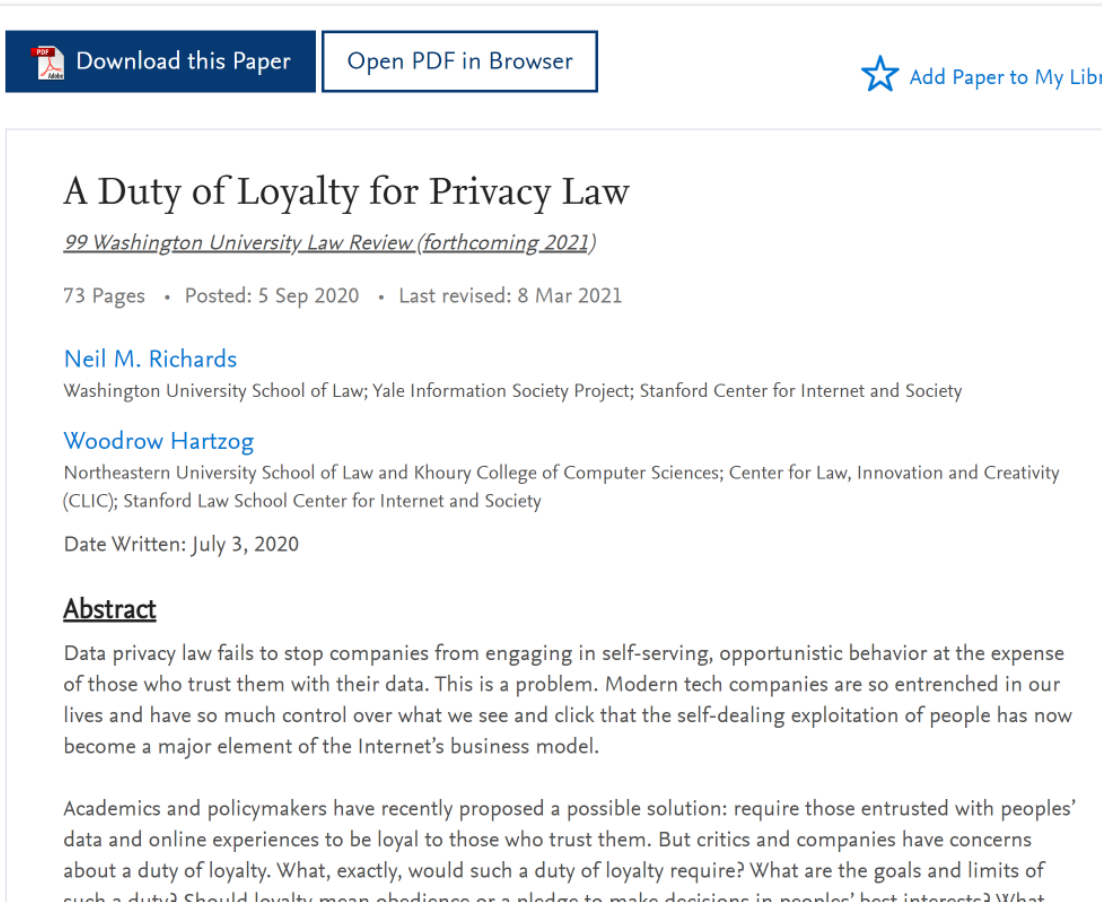
Woodrow Hartzog
Northeastern University School of Law and Khoury College of Computer Sciences; Center for Law, Information and Creativity (CLIC); Stanford Law School Center for Internet and Society

Date Written: September 3, 2015

Abstract

Trust is beautiful. The willingness to accept vulnerability to the actions of others is the essential ingredient for friendship, commerce, transportation, and virtually every other activity that involves other people. It allows us to build things, and it allows us to grow. Trust is everywhere, but particularly at the core of the information relationships that have come to characterize our modern, digital lives. Relationships between people and their ISPs, social networks, and hired professionals are typically understood in terms of privacy. But the way we have talked about privacy has a pessimism problem – privacy is conceptualized in negative terms, which leads us to mistakenly look for "creepy" new practices, focus excessively on harms from invasions of privacy, and place too much weight on the ability of individuals to opt out of harmful or offensive data practices.

LOYALTY



A Duty of Loyalty for Privacy Law

99 Washington University Law Review (forthcoming 2021)

73 Pages • Posted: 5 Sep 2020 • Last revised: 8 Mar 2021

Neil M. Richards
Washington University School of Law; Yale Information Society Project; Stanford Center for Internet and Society

Woodrow Hartzog
Northeastern University School of Law and Khoury College of Computer Sciences; Center for Law, Innovation and Creativity (CLIC); Stanford Law School Center for Internet and Society

Date Written: July 3, 2020

Abstract

Data privacy law fails to stop companies from engaging in self-serving, opportunistic behavior at the expense of those who trust them with their data. This is a problem. Modern tech companies are so entrenched in our lives and have so much control over what we see and click that the self-dealing exploitation of people has now become a major element of the Internet's business model.

Academics and policymakers have recently proposed a possible solution: require those entrusted with peoples' data and online experiences to be loyal to those who trust them. But critics and companies have concerns about a duty of loyalty. What, exactly, would such a duty of loyalty require? What are the goals and limits of such a duty? Should loyalty mean obedience or a pledge to make decisions in peoples' best interests? What

A Proposed Duty of Loyalty

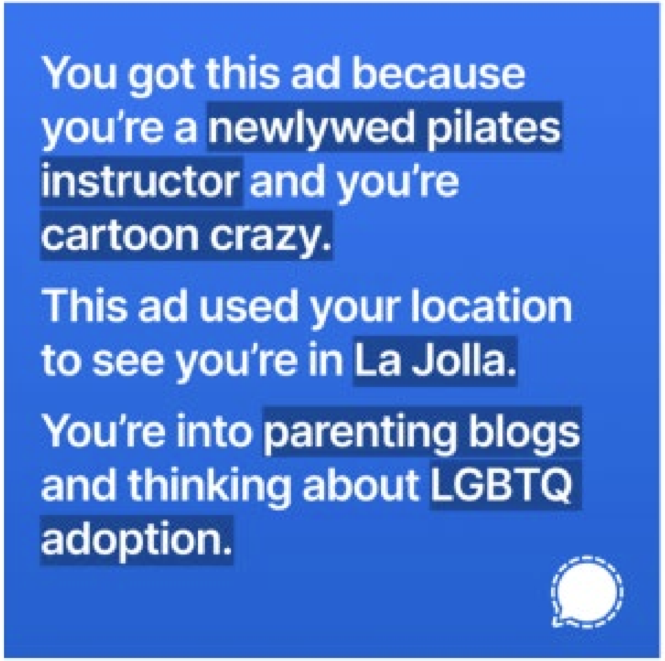
Trusted parties may not process personal data or design information technologies in a way that substantially conflicts with the best interests of a person with respect to—

(1) the experience of the person when using a platform owned or controlled by the trusted party;

or

(2) the personal data of the person

Loyal Personalization

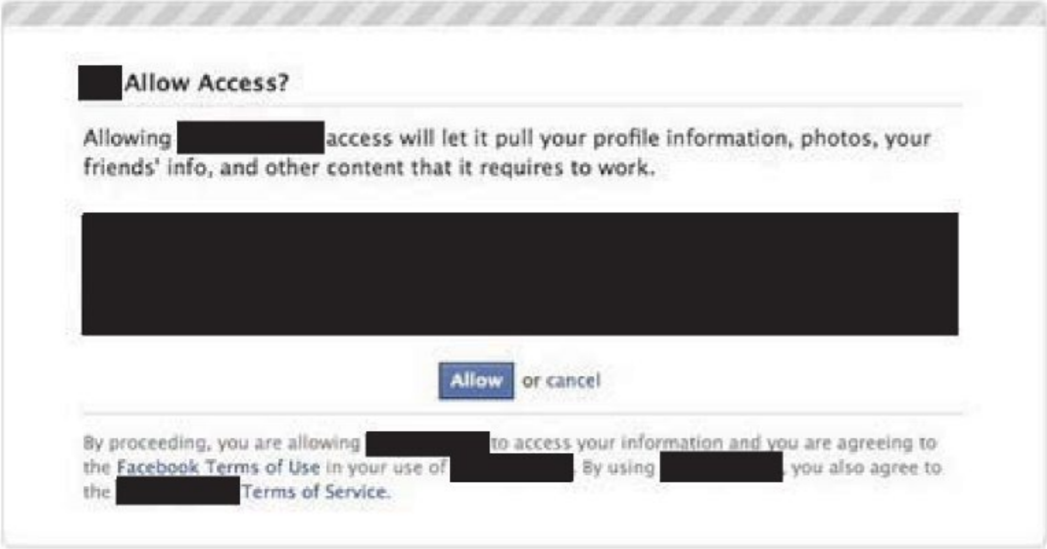


You got this ad because you're a newlywed pilates instructor and you're cartoon crazy.

This ad used your location to see you're in La Jolla.

You're into parenting blogs and thinking about LGBTQ adoption.

Loyal Gatekeeping



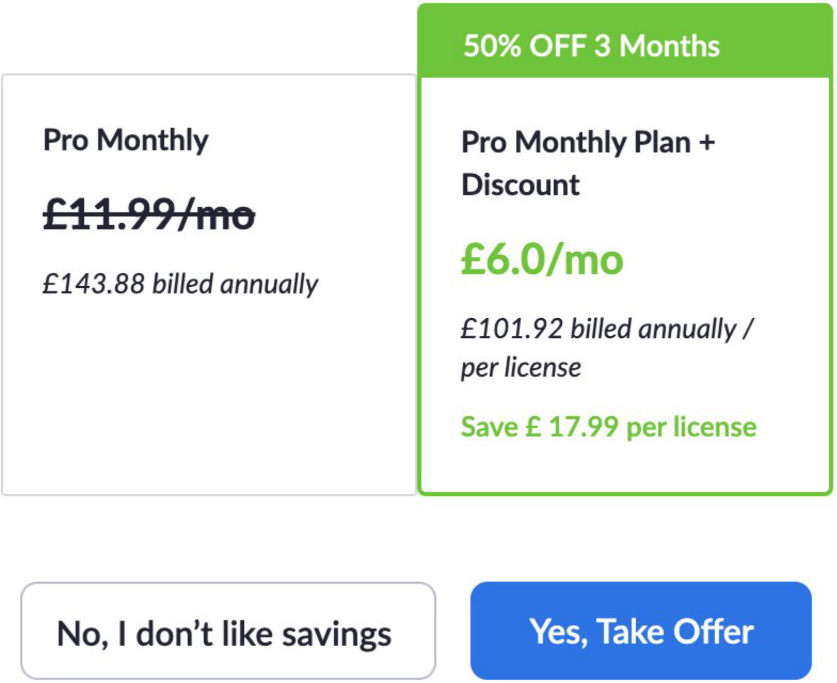
[redacted] Allow Access?

Allowing [redacted] access will let it pull your profile information, photos, your friends' info, and other content that it requires to work.

[redacted]

By proceeding, you are allowing [redacted] to access your information and you are agreeing to the Facebook Terms of Use in your use of [redacted]. By using [redacted] you also agree to the [redacted] Terms of Service.

Loyal Influencing



Pro Monthly

~~£11.99/mo~~

£143.88 billed annually

50% OFF 3 Months

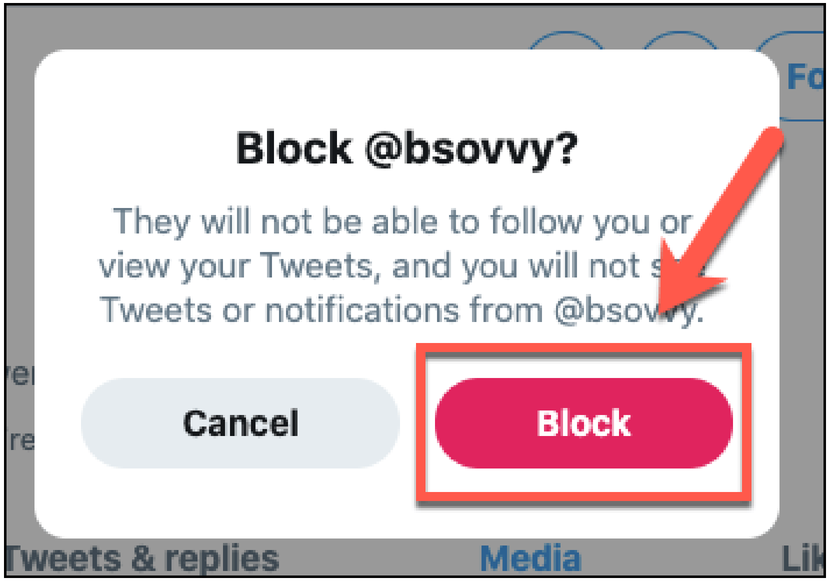
Pro Monthly Plan + Discount

£6.0/mo

£101.92 billed annually / per license

Save £ 17.99 per license

Loyal Mediation



Block @bsovvy?

They will not be able to follow you or view your Tweets, and you will not s Tweets or notifications from @bsovvy.

Conclusion

While loyalty is only one piece of the puzzle of making the best of our information revolution, it can be the key piece that makes all of the others work.

Robophobia in Medicine

Christopher Robertson

School of Law

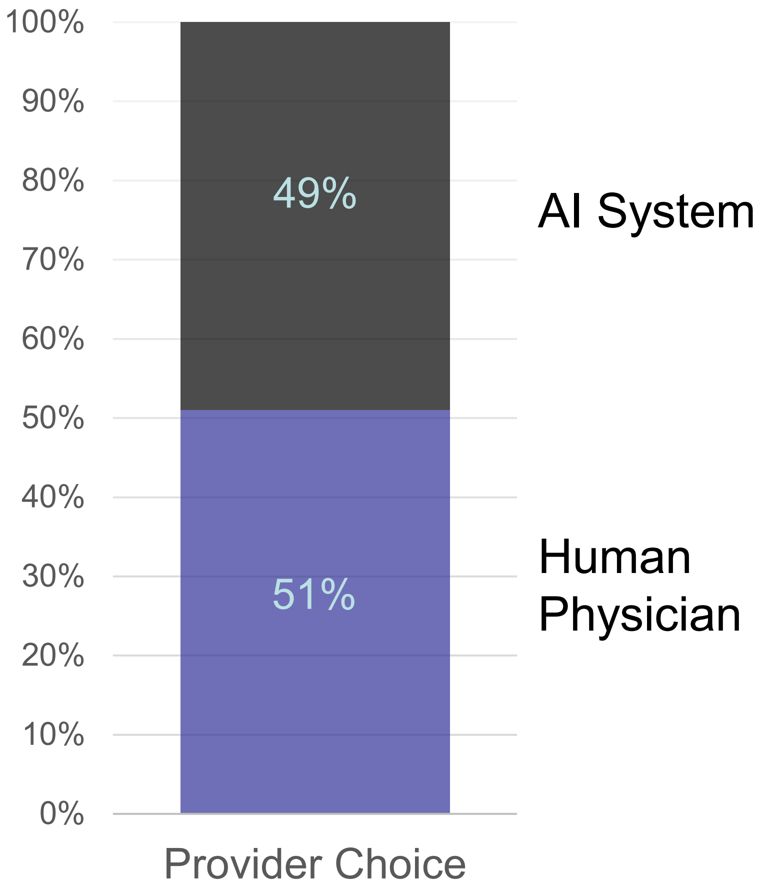
co-authored work, funded by NIH 3R25HL126140-05S1

What if you had a deadly disease …

to get a specific diagnosis and a treatment plan, would you choose:

- a specialist human physician or
- a specialized AI-system?

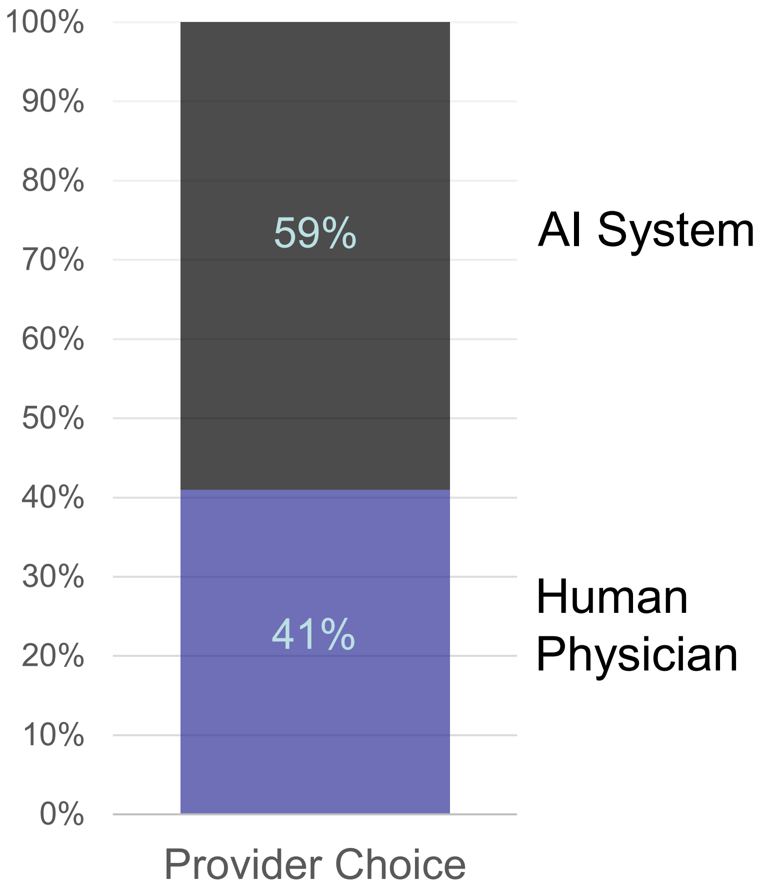
Your insurance covers either one.



n = 2675, nationally representative sample, MOE +/- 3%

What if you had a deadly disease …

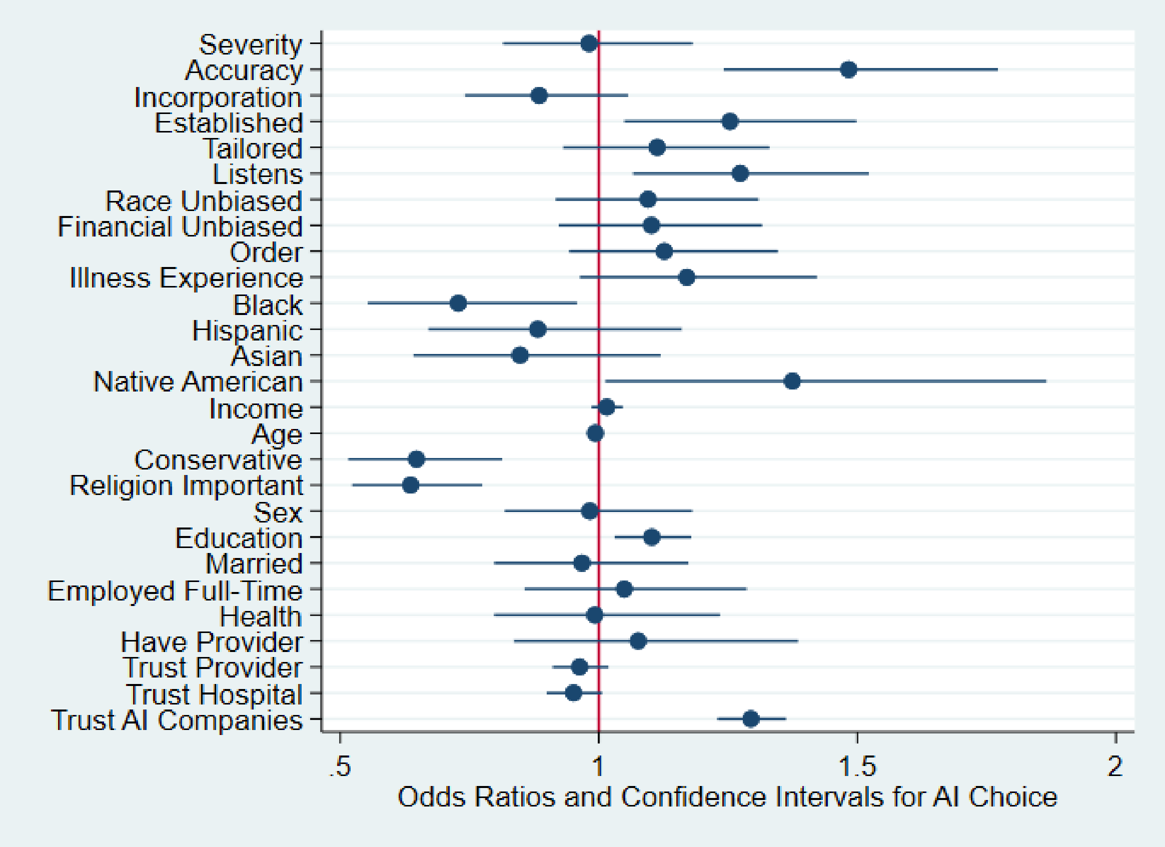
"Your doctor tells you that, based on scientific studies in leading journals, **the AI system is proven more accurate** at diagnoses compared to even specialist human physicians."



n = 2675, nationally representative sample, MOE +/- 3%

AI Aversion is…

- **A health risk.**
- Robust.
- Especially strong for conservative, religious, older, and Black Americans.

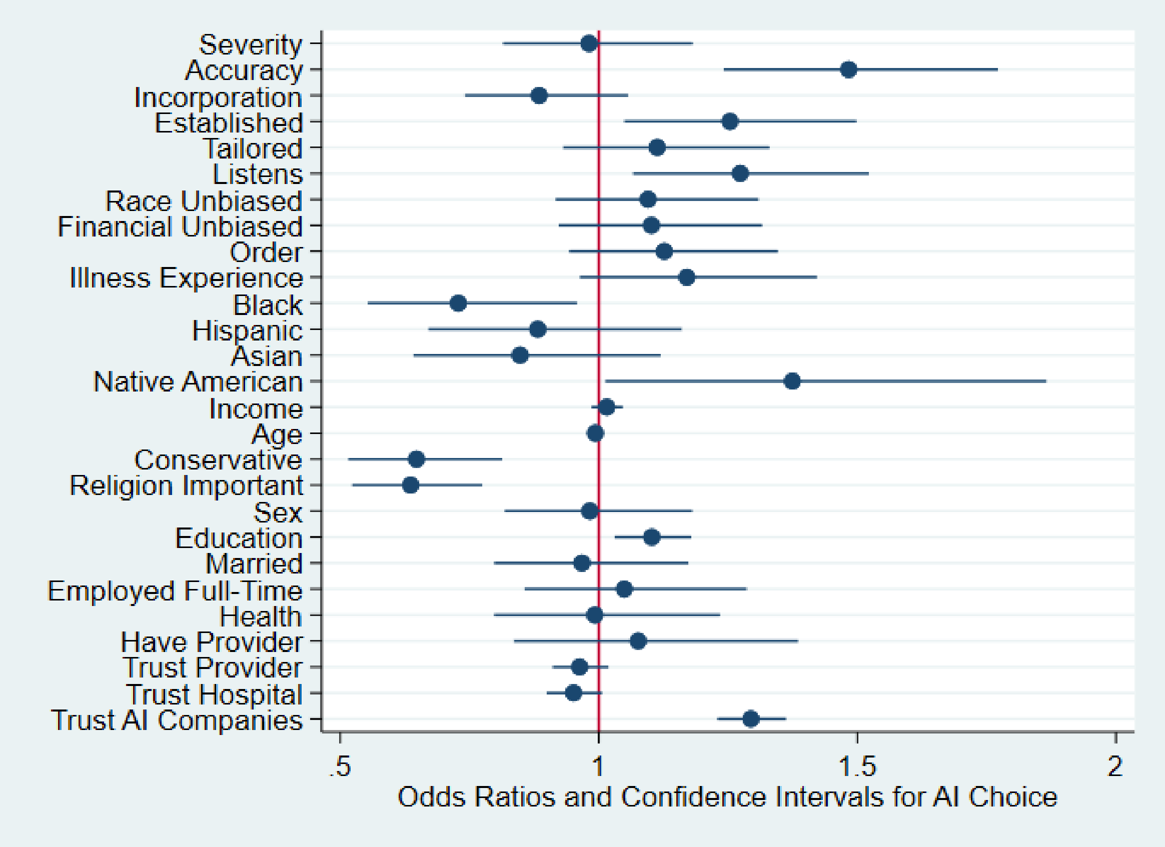


n = 2675

AI Aversion can be mitigated by…

- Education.
- Nudges.
- A listening user-experience.
- Public trust in AI companies.

But many patients will still decline AI, if given the choice.



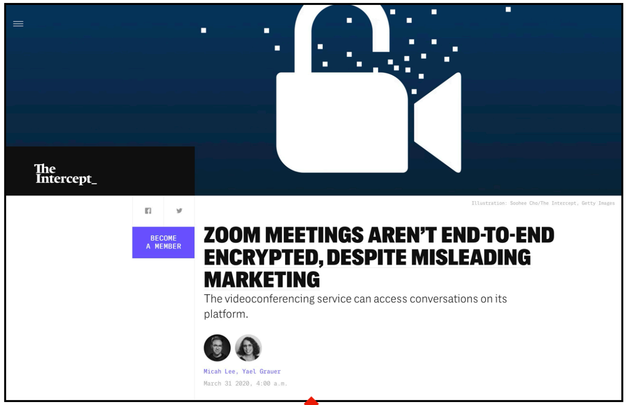
n = 2675

Independent Research as a Necessary Input to Software Accountability

Andy Sellars

Clinical Associate Professor
Executive Director, BU/MIT Student Innovations Law Clinic
School of Law

"We use end-to-end encryption"



The Intercept_

Illustration: Soohee Cho/The Intercept, Getty Images

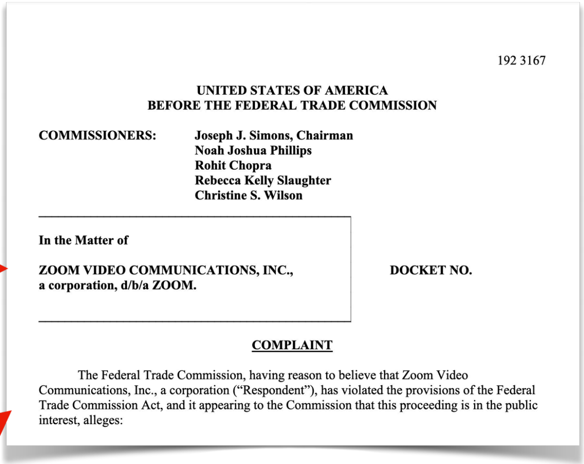
BECOME A MEMBER

ZOOM MEETINGS AREN'T END-TO-END ENCRYPTED, DESPITE MISLEADING MARKETING

The videoconferencing service can access conversations on its platform.

Micah Lee, Yael Grauer

March 31 2020, 4:00 a.m.



192 3167

UNITED STATES OF AMERICA
BEFORE THE FEDERAL TRADE COMMISSION

COMMISSIONERS: Joseph J. Simons, Chairman
Noah Joshua Phillips
Rohit Chopra
Rebecca Kelly Slaughter
Christine S. Wilson

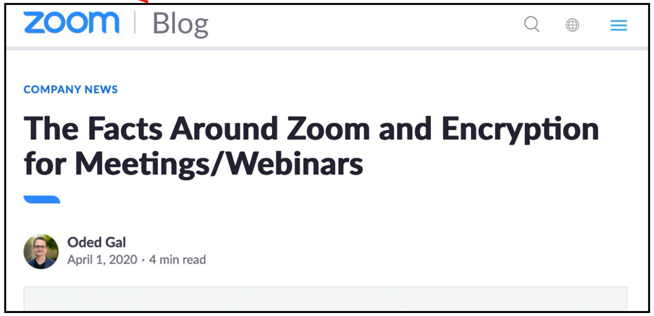
In the Matter of

ZOOM VIDEO COMMUNICATIONS, INC.,
a corporation, d/b/a ZOOM.

DOCKET NO.

COMPLAINT

The Federal Trade Commission, having reason to believe that Zoom Video Communications, Inc., a corporation ("Respondent"), has violated the provisions of the Federal Trade Commission Act, and it appearing to the Commission that this proceeding is in the public interest, alleges:

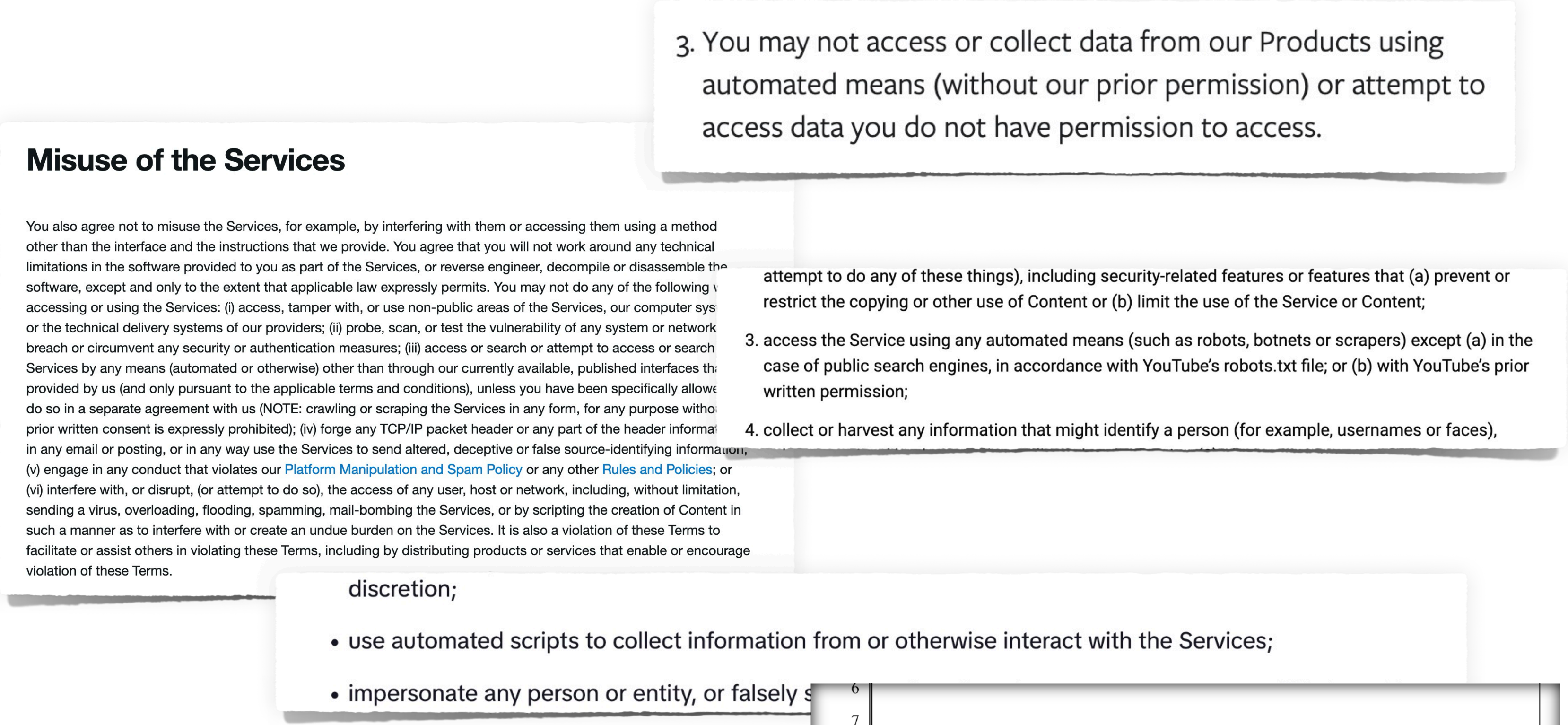


zoom | Blog

COMPANY NEWS

The Facts Around Zoom and Encryption for Meetings/Webinars

Oded Gal
April 1, 2020 · 4 min read



3. You may not access or collect data from our Products using automated means (without our prior permission) or attempt to access data you do not have permission to access.

Misuse of the Services

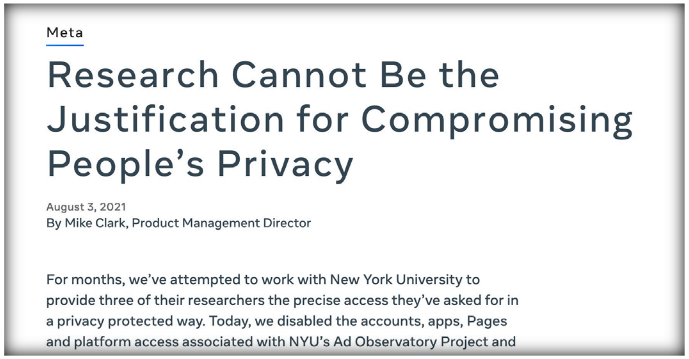
You also agree not to misuse the Services, for example, by interfering with them or accessing them using a method other than the interface and the instructions that we provide. You agree that you will not work around any technical limitations in the software provided to you as part of the Services, or reverse engineer, decompile or disassemble the software, except and only to the extent that applicable law expressly permits. You may not do any of the following v accessing or using the Services: (i) access, tamper with, or use non-public areas of the Services, our computer sys or the technical delivery systems of our providers; (ii) probe, scan, or test the vulnerability of any system or network breach or circumvent any security or authentication measures; (iii) access or search or attempt to access or search Services by any means (automated or otherwise) other than through our currently available, published interfaces th provided by us (and only pursuant to the applicable terms and conditions), unless you have been specifically allowe do so in a separate agreement with us (NOTE: crawling or scraping the Services in any form, for any purpose witho prior written consent is expressly prohibited); (iv) forge any TCP/IP packet header or any part of the header informat in any email or posting, or in any way use the Services to send altered, deceptive or false source-identifying information, (v) engage in any conduct that violates our Platform Manipulation and Spam Policy or any other Rules and Policies; or (vi) interfere with, or disrupt, (or attempt to do so), the access of any user, host or network, including, without limitation, sending a virus, overloading, flooding, spamming, mail-bombing the Services, or by scripting the creation of Content in such a manner as to interfere with or create an undue burden on the Services. It is also a violation of these Terms to facilitate or assist others in violating these Terms, including by distributing products or services that enable or encourage violation of these Terms.

attempt to do any of these things), including security-related features or features that (a) prevent or restrict the copying or other use of Content or (b) limit the use of the Service or Content;

3. access the Service using any automated means (such as robots, botnets or scrapers) except (a) in the case of public search engines, in accordance with YouTube's robots.txt file; or (b) with YouTube's prior written permission;
4. collect or harvest any information that might identify a person (for example, usernames or faces),

discretion;

- use automated scripts to collect information from or otherwise interact with the Services;
- impersonate any person or entity, or falsely s

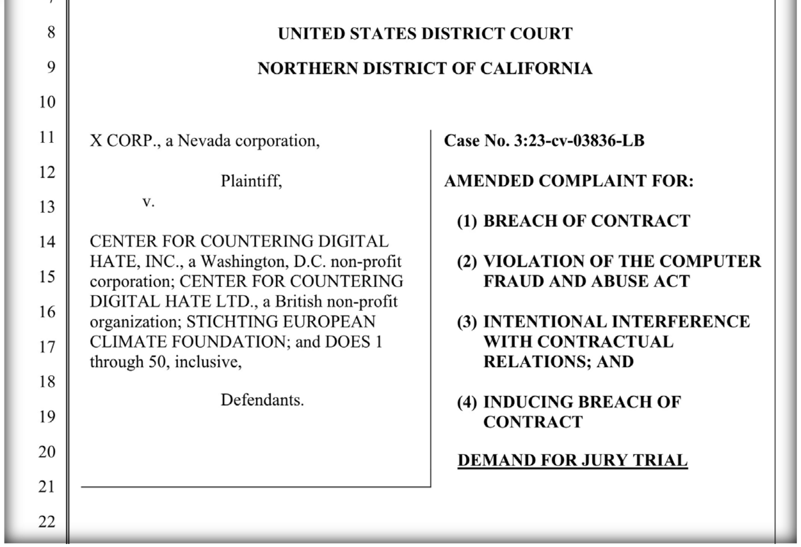


Meta

Research Cannot Be the Justification for Compromising People's Privacy

August 3, 2021
By Mike Clark, Product Management Director

For months, we've attempted to work with New York University to provide three of their researchers the precise access they've asked for in a privacy protected way. Today, we disabled the accounts, apps, Pages and platform access associated with NYU's Ad Observatory Project and



UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

X CORP., a Nevada corporation,

Plaintiff,

v.

CENTER FOR COUNTERING DIGITAL HATE, INC., a Washington, D.C. non-profit corporation; CENTER FOR COUNTERING DIGITAL HATE LTD., a British non-profit organization; STICHTING EUROPEAN CLIMATE FOUNDATION; and DOES 1 through 50, inclusive,

Defendants.

Case No. 3:23-cv-03836-LB

AMENDED COMPLAINT FOR:

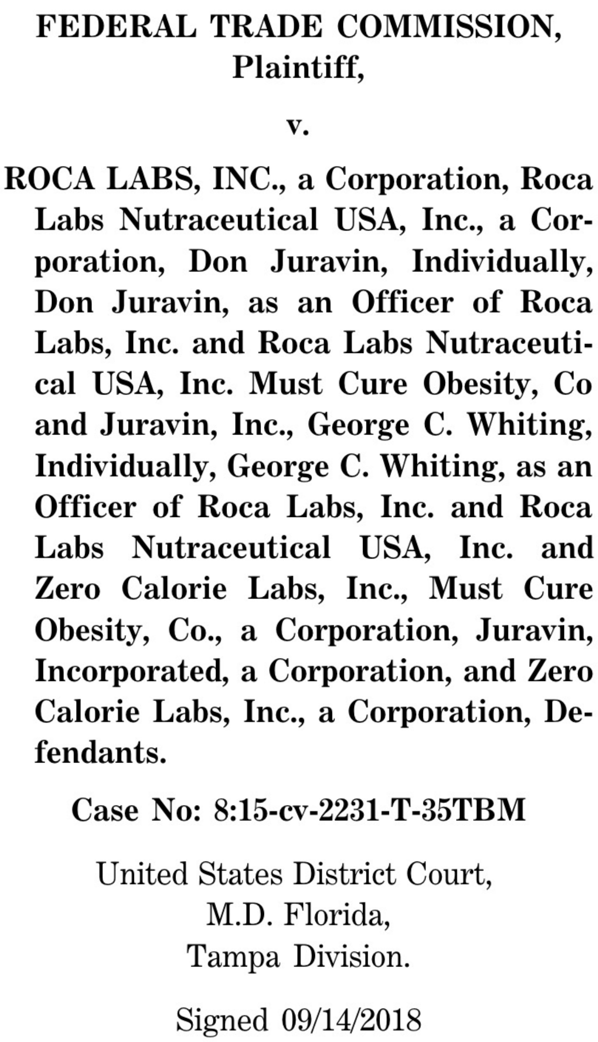
(1) BREACH OF CONTRACT

(2) VIOLATION OF THE COMPUTER FRAUD AND ABUSE ACT

(3) INTENTIONAL INTERFERENCE WITH CONTRACTUAL RELATIONS; AND

(4) INDUCING BREACH OF CONTRACT

DEMAND FOR JURY TRIAL



FEDERAL TRADE COMMISSION,
Plaintiff,

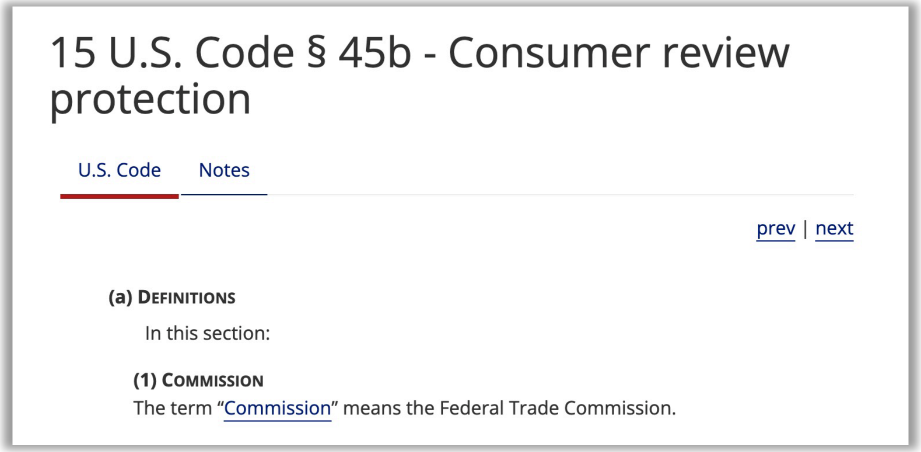
v.

ROCA LABS, INC., a Corporation, Roca Labs Nutraceutical USA, Inc., a Corporation, Don Juravin, Individually, Don Juravin, as an Officer of Roca Labs, Inc. and Roca Labs Nutraceutical USA, Inc. Must Cure Obesity, Co and Juravin, Inc., George C. Whiting, Individually, George C. Whiting, as an Officer of Roca Labs, Inc. and Roca Labs Nutraceutical USA, Inc. and Zero Calorie Labs, Inc., Must Cure Obesity, Co., a Corporation, Juravin, Incorporated, a Corporation, and Zero Calorie Labs, Inc., a Corporation, Defendants.

Case No: 8:15-cv-2231-T-35TBM

United States District Court,
M.D. Florida,
Tampa Division.

Signed 09/14/2018



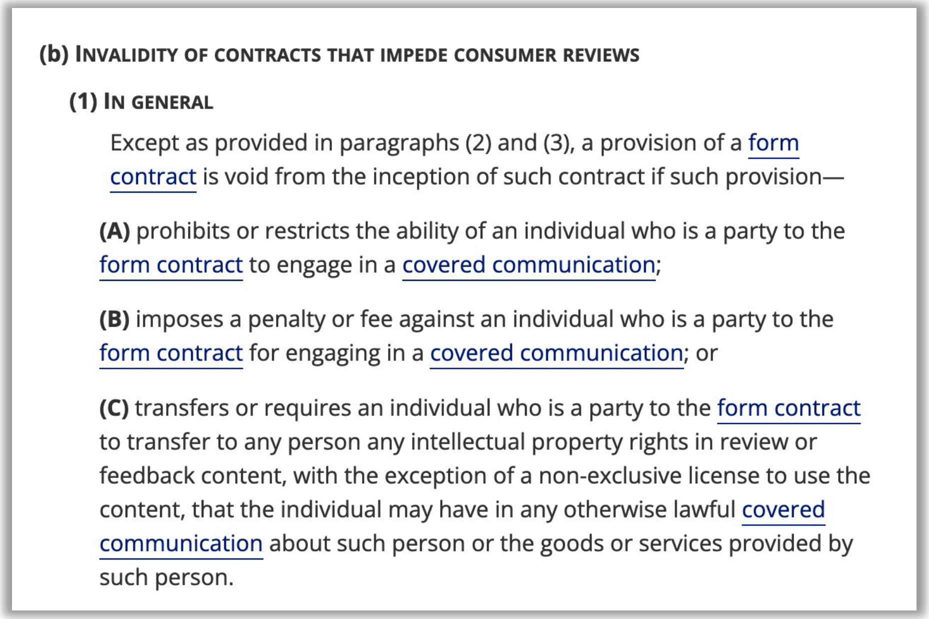
15 U.S. Code § 45b - Consumer review protection

(a) Definitions

In this section:

(1) Commission

The term "Commission" means the Federal Trade Commission.



(b) Invalidity of contracts that impede consumer reviews

(1) In general

Except as provided in paragraphs (2) and (3), a provision of a form contract is void from the inception of such contract if such provision—

(A) prohibits or restricts the ability of an individual who is a party to the form contract to engage in a covered communication;

(B) imposes a penalty or fee against an individual who is a party to the form contract for engaging in a covered communication; or

(C) transfers or requires an individual who is a party to the form contract to transfer to any person any intellectual property rights in review or feedback content, with the exception of a non-exclusive license to use the content, that the individual may have in any otherwise lawful covered communication about such person or the goods or services provided by such person.

We Need to Audit Algorithms

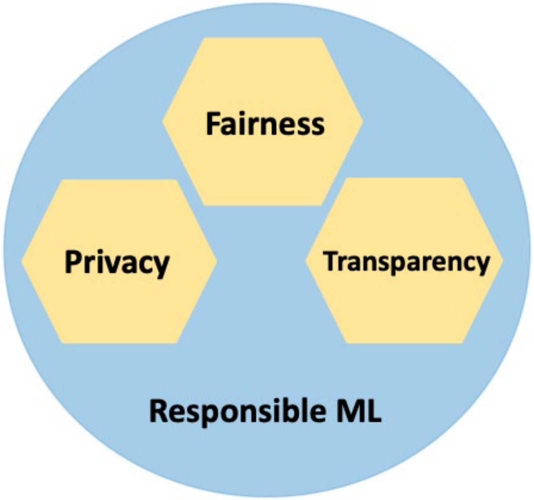
Mark Crovella

Professor, CAS Computer Science
Professor, Faculty of Computing and Data Sciences
Chair of Academic Affairs, CDS

The Need for Algorithmic Audits

- Machine learning algorithms are increasingly used in ways that affect individual lives and have social impacts
- COMPAS used by US courts to assess likelihood of a defendant becoming a recidivist
- Algorithms can inherit biases from training data, or from the goals designed in
- Modern machine learning algorithms are not ***transparent*** – we cannot easily understand why they make the decisions they do
- Hence:
  - We must "audit" algorithms
  - Usually, we do this from the ***outside***

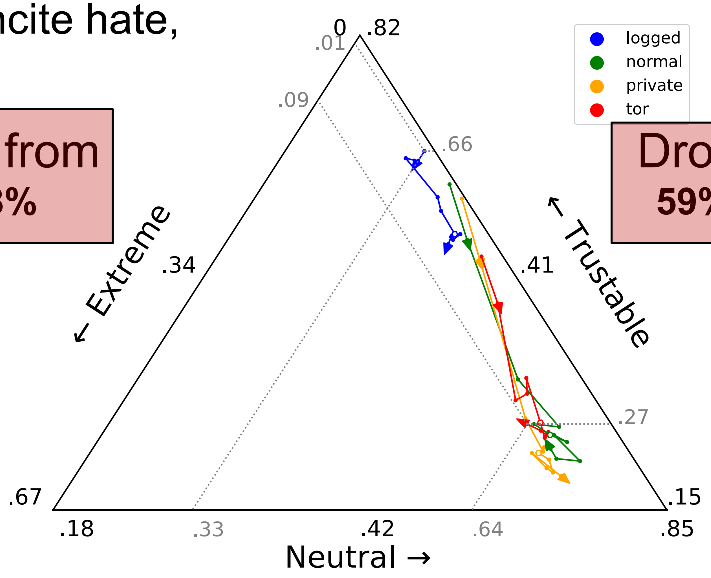
- We are engaged in efforts to
  - Perform audits
  - Design auditing algorithms
- To improve the social impacts of machine learning



Behavioral Auditing: YouTube's Garden Path

- Each video watched on YouTube comes with a set of recommendations to be viewed next
- Where do these recommendations lead?
- In general, to more extreme and less trustworthy content
- Extreme channels deny established knowledge, incite hate, or promote fake news

- Most of the change occurs within 5 "clicks"
- Biggest surprise: users with strong privacy settings (anonymous browsing) see much greater effect!



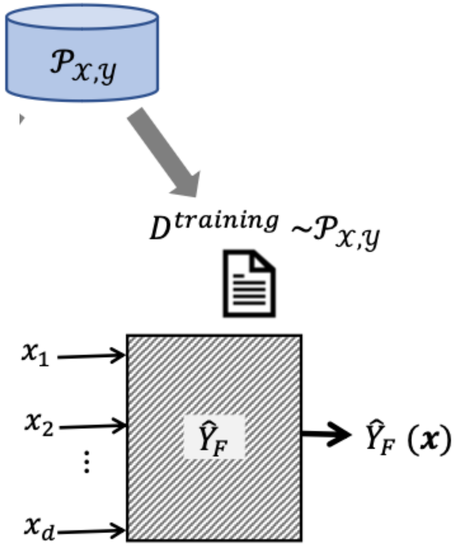
Increases from **1.2%** to **8%**

Drops from **59%** to **31%**

Data Usage Auditing: *Data Minimization*

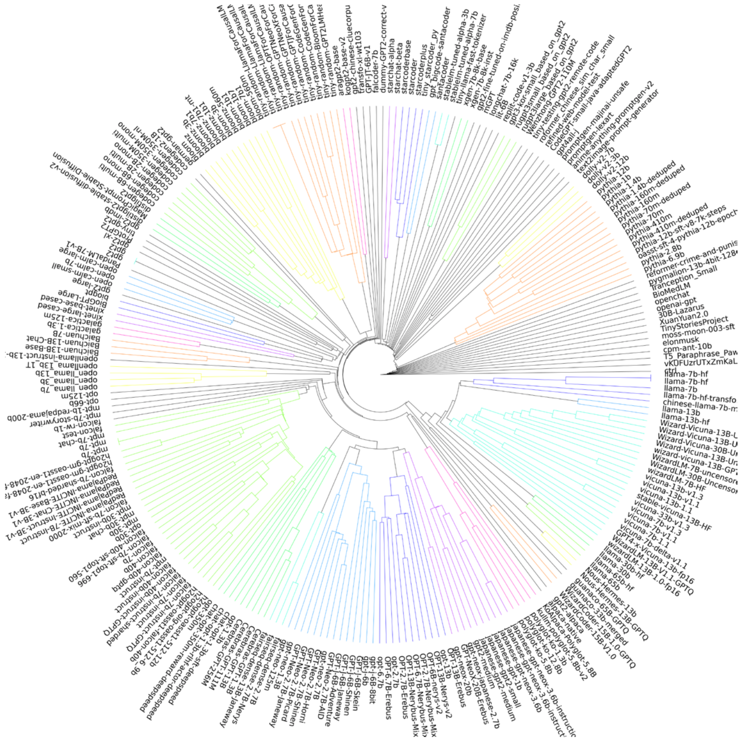
- Machine Learning models are data-hungry!
- But what is "need-to-know" for an algorithm?
- Should you need to disclose highly personal information to obtain a bank loan?
- How can we define and detect this sort of privacy violation?
- GDPR: "Personal data shall be … limited to what is necessary … to the purposes for which they are processed"
- We operationalize this concept in terms of a particular kind of audit
- We ask whether the system's decision is affected by the value of each input

- We define an auditing algorithm to efficiently detect when data is *not needed* by the algorithm making a decision


$\mathcal{P}_{x,y}$
$D^{training} \sim \mathcal{P}_{x,y}$
x_1
x_2
⋮
x_d
$\hat{Y}_F$
$\hat{Y}_F\ (\boldsymbol{x})$

Auditing Language Models: Beyond the Benchmarks

- Large Language Models like ChatGPT are poised to dramatically change the way we work and learn
- ChatGPT reached 100 million users just two months after its launch
- LLMs are and will become highly specialized and numerous
- Currently there are more than 300,000 LLMs on HuggingFace
- Current methods for evaluating LLMs focus on accuracy on benchmarks
- Psychometrics: methods for evaluating both tests and test-takers simultaneously
- Our goal: better tools for determining the nature of any given LLM and its fitness to purpose



Evolutionary Tree of LLMs
Source https://arxiv.org/abs/2307.09793

Choice Architecture, Privacy Valuations, and Selection Bias in Consumer Data

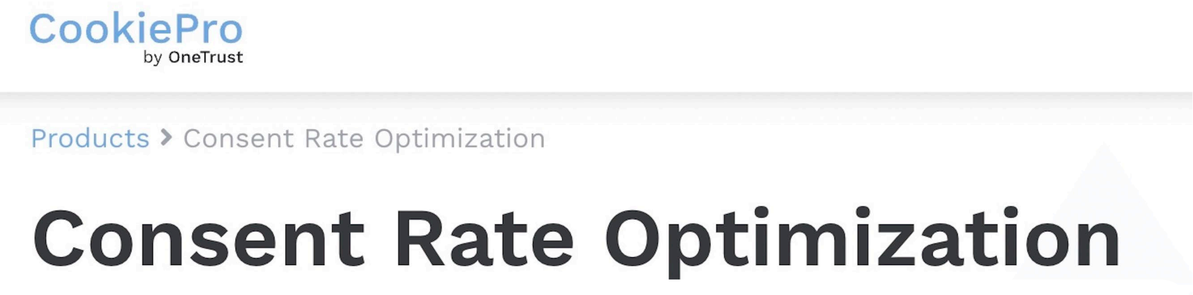
Tesary Lin

Assistant Professor, Questrom

We use cookies to enhance your experience on our site, analyze site usage, and assist in our marketing activities. You can manage your preferences in Cookie Settings. By clicking Accept Cookies or closing this dialogue you agree to our use of cookies. If you do not allow cookies, then some features of our site may not function properly. To learn more, visit our cookie notice.

Cookie Settings ACCEPT COOKIES ×

Biased sample → Biased algorithms → Biased decisions

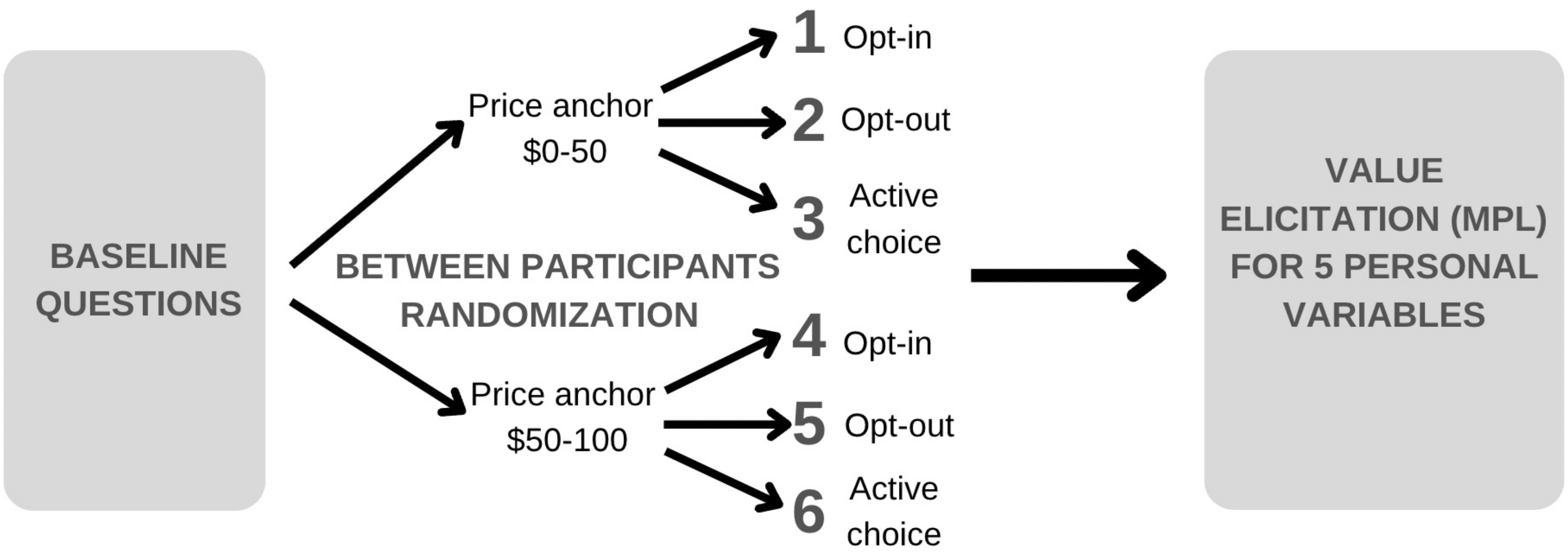


CookiePro
by OneTrust

Products › Consent Rate Optimization

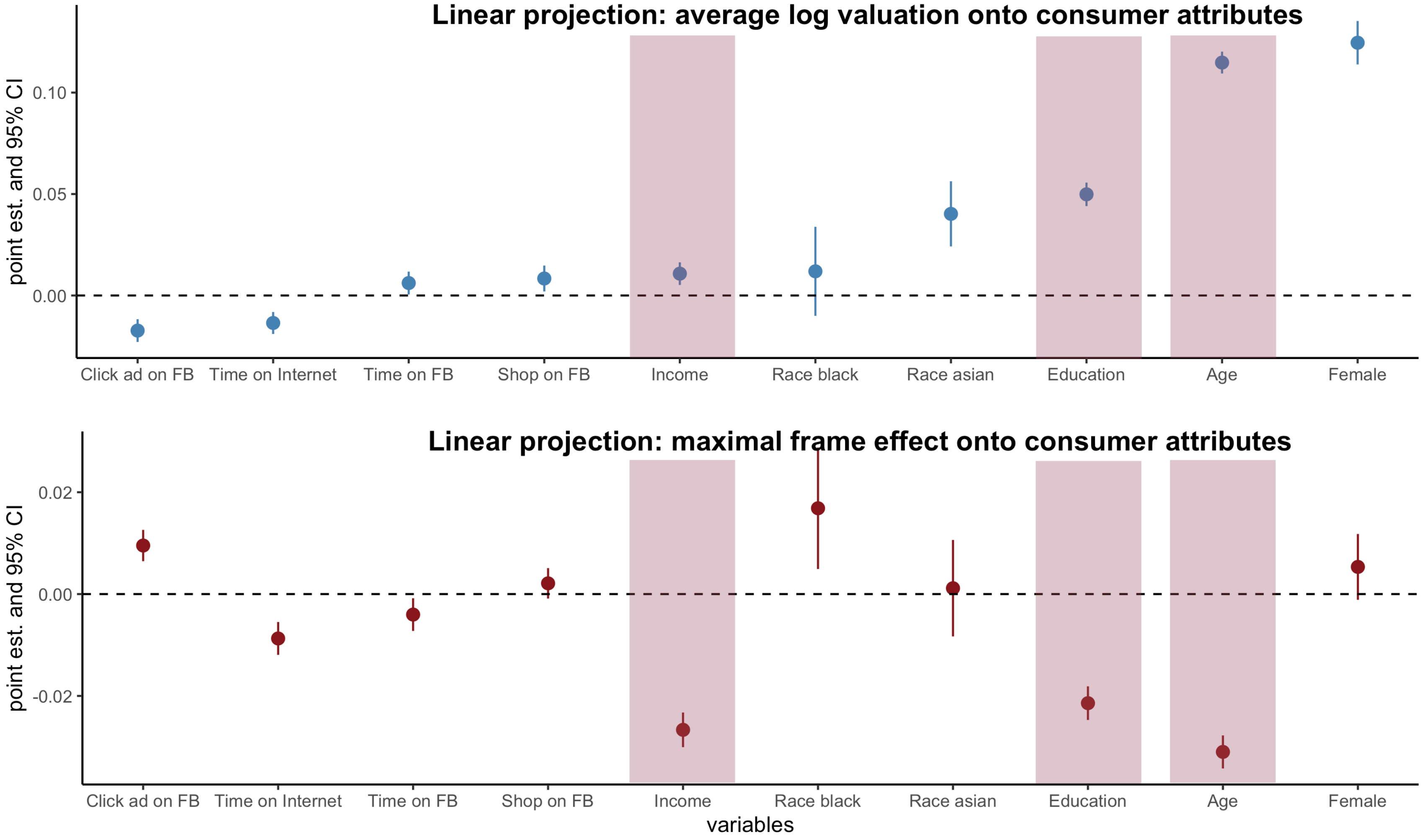
Consent Rate Optimization

How does choice architecture affect sample bias?
---We run an experiment to measure its causal effects

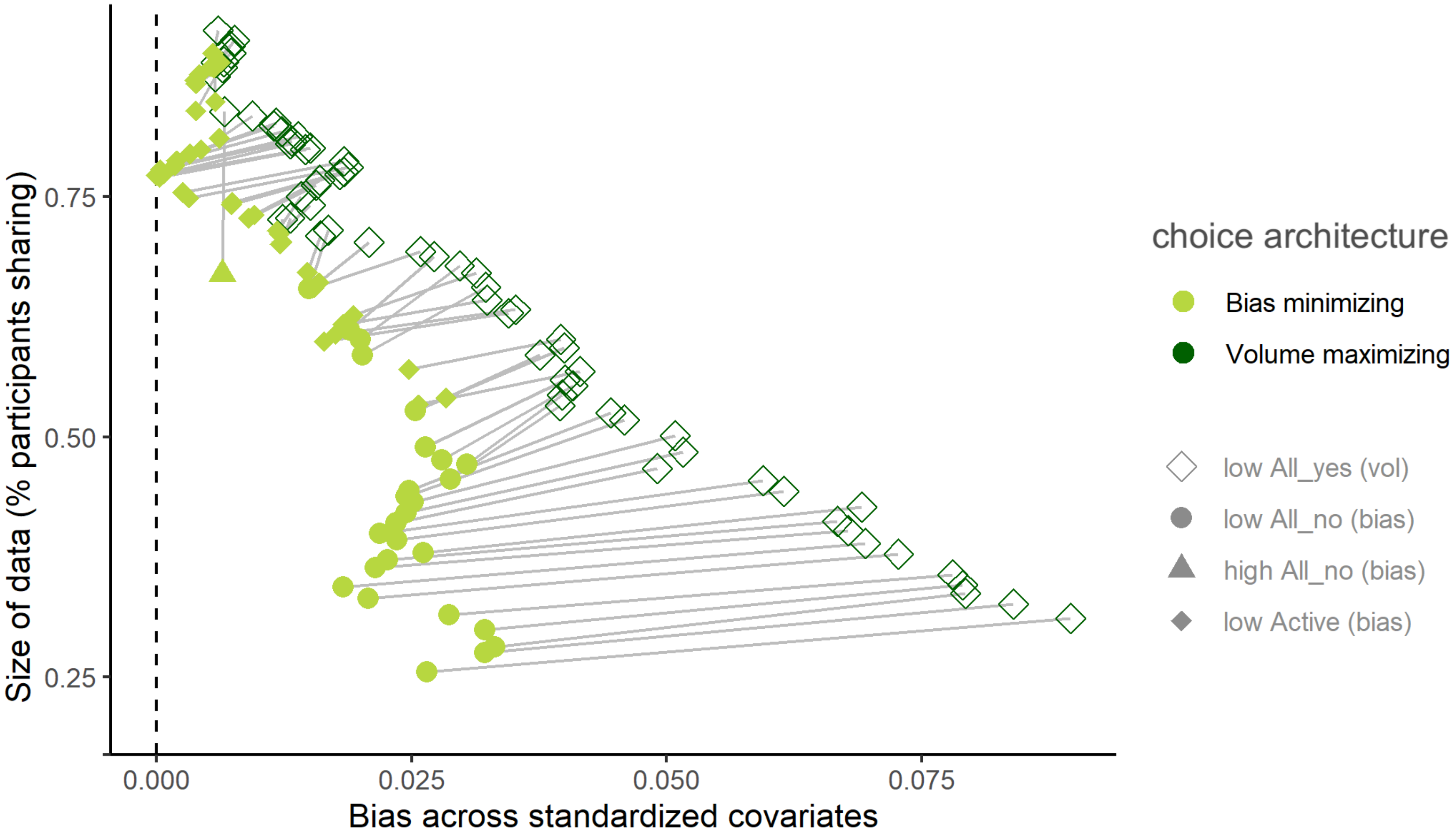


5k participants; willingness-to-accept to share FB data with advertisers

Younger, Poorer, Less Educated consumers value privacy less & are more easily nudged by choice architecture



Choice architecture that maximizes sample volume → potential for a more biased sample



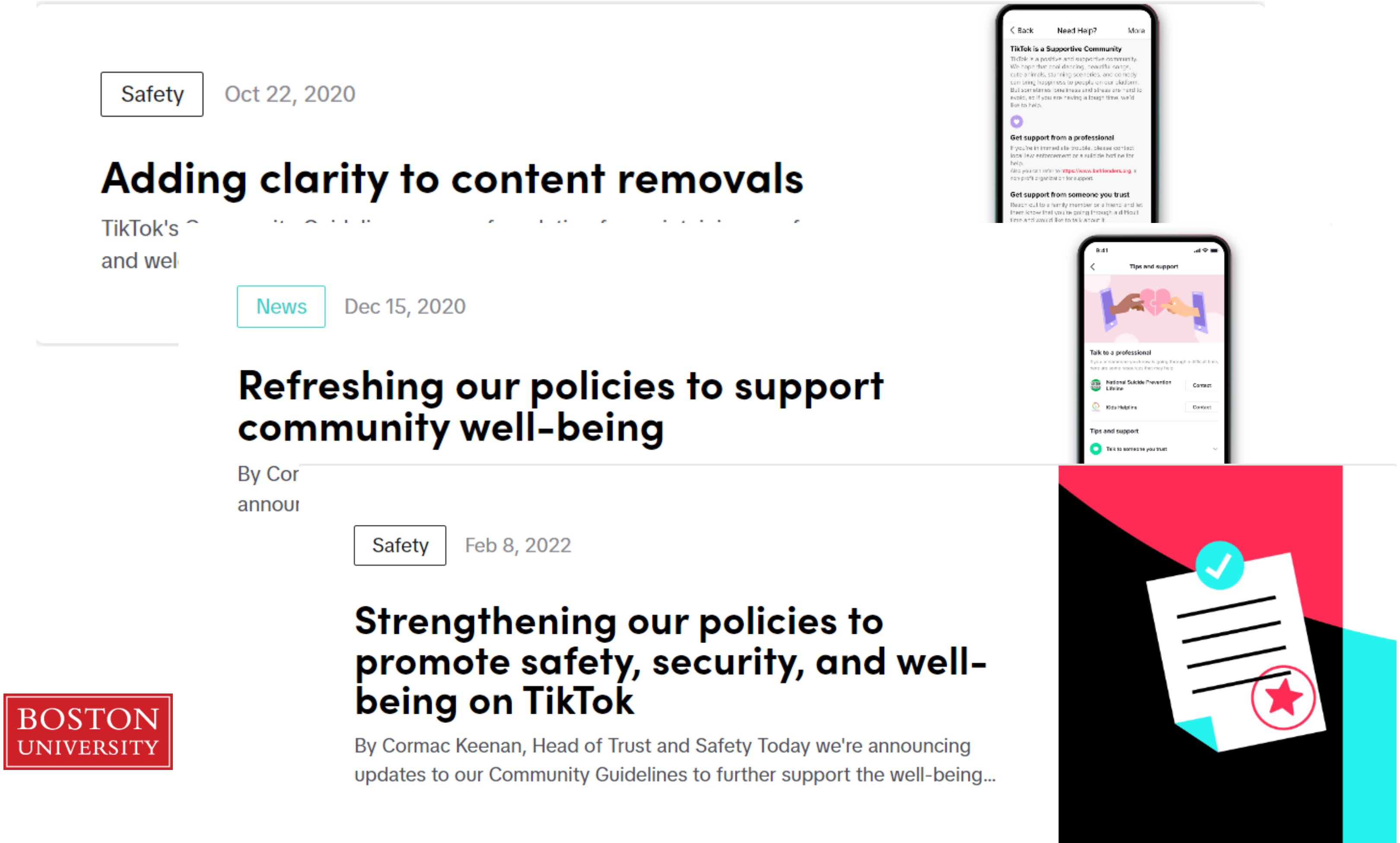
Shifting Platform Values in Community Guidelines: The Evolution of Governance Frameworks

Chris Chao Su

Assistant Professor of Emerging Media Studies
College of Communication + Computing & Data Science

Community Guidelines as *Discursive Performance* and *Governance Frameworks*

Community Guidelines



Safety Oct 22, 2020

Adding clarity to content removals

TikTok's

and wel

< Back Need Help? More

TikTok is a Supportive Community

TikTok is a positive and supportive community. We hope that cool dancing, beautiful songs, cute animals, stunning sceneries, and comedy can bring happiness to people on our platform. But sometimes loneliness and stress are hard to avoid, so if you are having a tough time, we'd like to help.

Get support from a professional

If you're in immediate trouble, please contact local law enforcement or a suicide hotline for help.
Also you can refer to https://www.befrienders.org, a non-profit organization for support.

Get support from someone you trust

Reach out to a family member or a friend and let them know that you're going through a difficult time and would like to talk about it.

News Dec 15, 2020

Refreshing our policies to support community well-being

By Cor

annour

9:41

Tips and support

Talk to a professional

If you or someone you know is going through a difficult time, here are some resources that may help

National Suicide Prevention Lifeline Contact

Kids Helpline Contact

Tips and support

Talk to someone you trust

Safety Feb 8, 2022

Strengthening our policies to promote safety, security, and well-being on TikTok

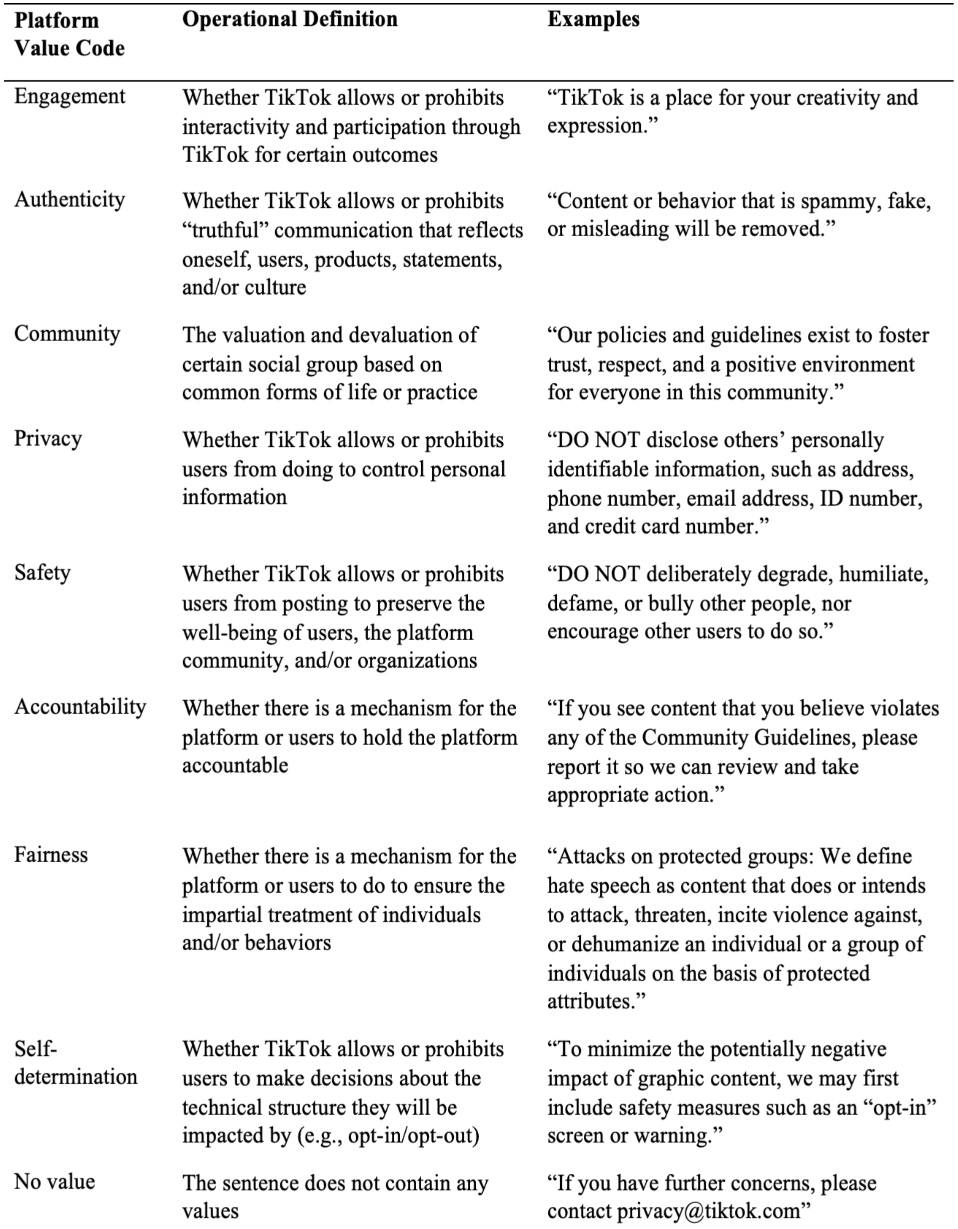
By Cormac Keenan, Head of Trust and Safety Today we're announcing updates to our Community Guidelines to further support the well-being...

Platform Values

- Values as the **ideals** expressed by a particular social entity, which may guide *subsequent actions and judgments* (Hallinan et al. 2022)

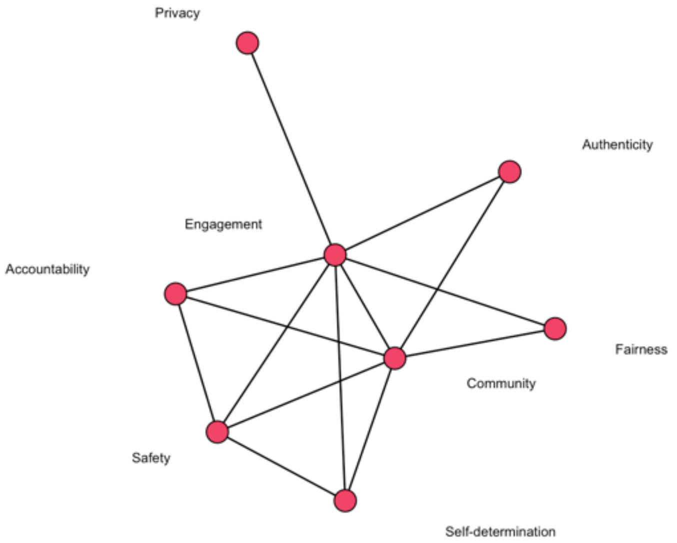
- Articulating the ideals about "how people ***should*** express themselves and interact with others" (Scharlach et al., 2023).

- Values as objects and principles

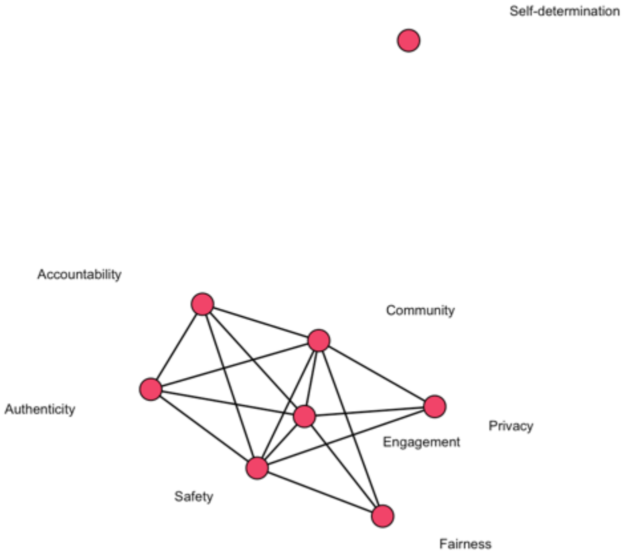


Platform Value Code	Operational Definition	Examples
Engagement	Whether TikTok allows or prohibits interactivity and participation through TikTok for certain outcomes	"TikTok is a place for your creativity and expression."
Authenticity	Whether TikTok allows or prohibits "truthful" communication that reflects oneself, users, products, statements, and/or culture	"Content or behavior that is spammy, fake, or misleading will be removed."
Community	The valuation and devaluation of certain social group based on common forms of life or practice	"Our policies and guidelines exist to foster trust, respect, and a positive environment for everyone in this community."
Privacy	Whether TikTok allows or prohibits users from doing to control personal information	"DO NOT disclose others' personally identifiable information, such as address, phone number, email address, ID number, and credit card number."
Safety	Whether TikTok allows or prohibits users from posting to preserve the well-being of users, the platform community, and/or organizations	"DO NOT deliberately degrade, humiliate, defame, or bully other people, nor encourage other users to do so."
Accountability	Whether there is a mechanism for the platform or users to hold the platform accountable	"If you see content that you believe violates any of the Community Guidelines, please report it so we can review and take appropriate action."
Fairness	Whether there is a mechanism for the platform or users to do to ensure the impartial treatment of individuals and/or behaviors	"Attacks on protected groups: We define hate speech as content that does or intends to attack, threaten, incite violence against, or dehumanize an individual or a group of individuals on the basis of protected attributes."
Self-determination	Whether TikTok allows or prohibits users to make decisions about the technical structure they will be impacted by (e.g., opt-in/opt-out)	"To minimize the potentially negative impact of graphic content, we may first include safety measures such as an "opt-in" screen or warning."
No value	The sentence does not contain any values	"If you have further concerns, please contact privacy@tiktok.com"

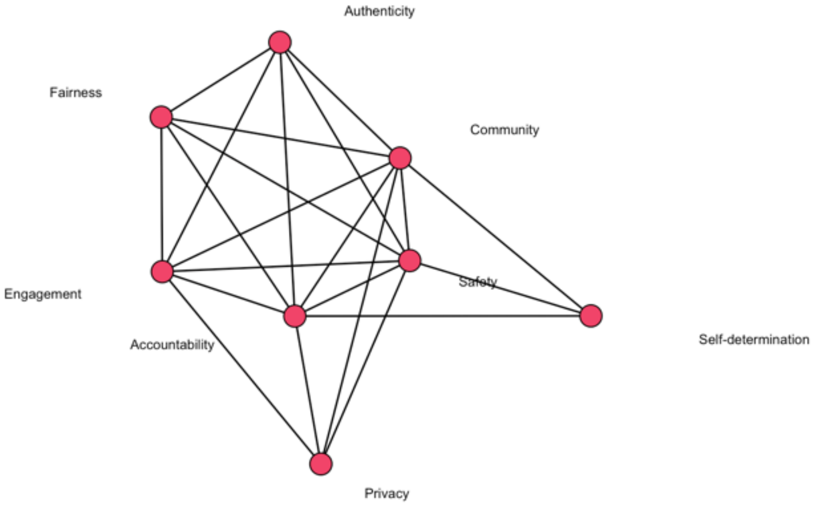
December 10, 2018



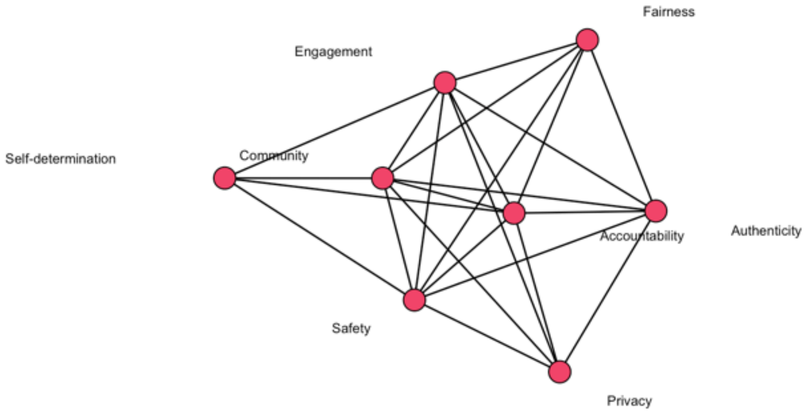
January 8, 2020

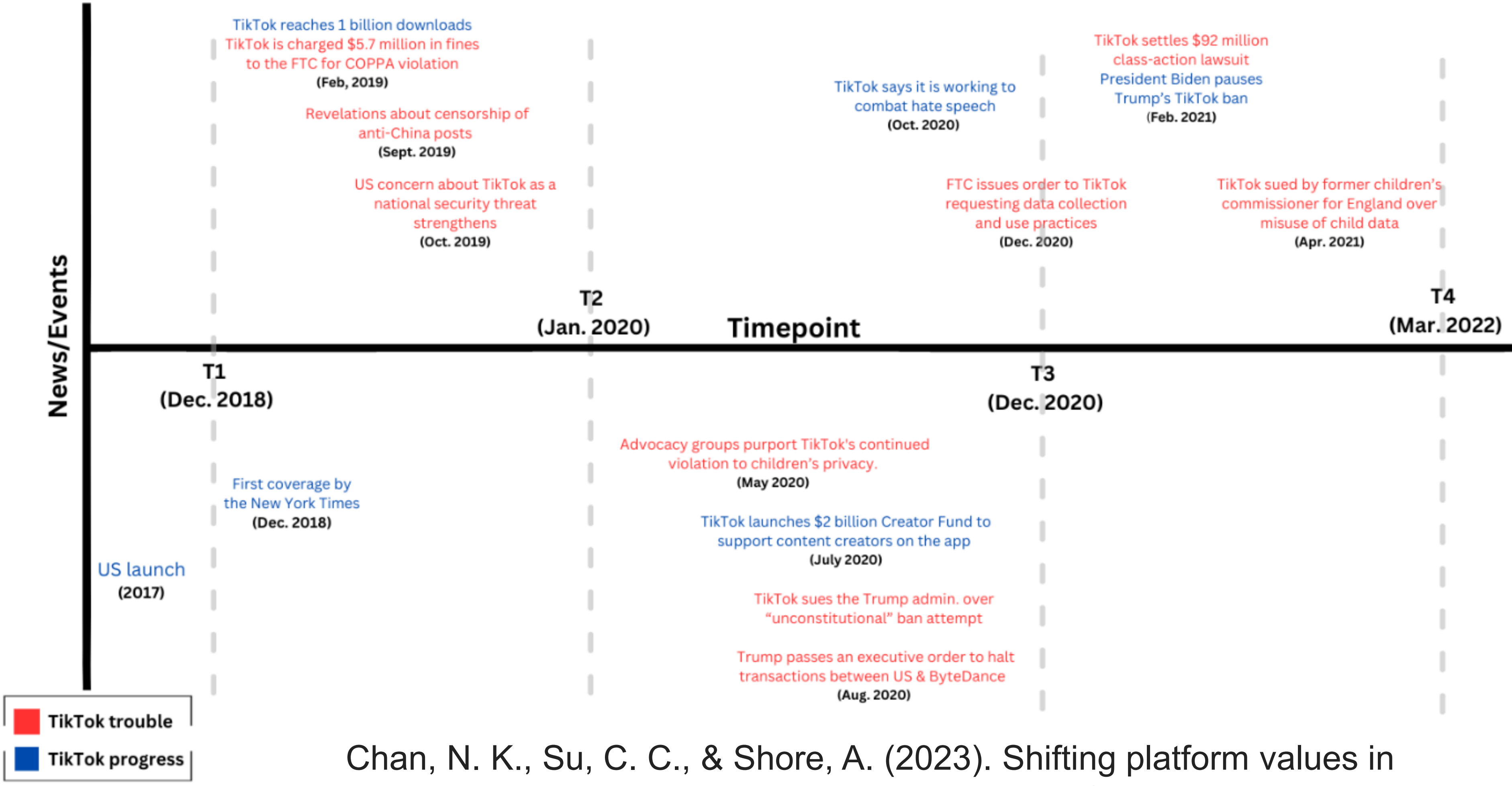


December 15, 2020



March 7, 2022





News/Events

Timepoint

T1 (Dec. 2018)

T2 (Jan. 2020)

T3 (Dec. 2020)

T4 (Mar. 2022)

US launch
(2017)

First coverage by the New York Times
(Dec. 2018)

TikTok reaches 1 billion downloads
TikTok is charged $5.7 million in fines to the FTC for COPPA violation
(Feb, 2019)

Revelations about censorship of anti-China posts
(Sept. 2019)

US concern about TikTok as a national security threat strengthens
(Oct. 2019)

Advocacy groups purport TikTok's continued violation to children's privacy.
(May 2020)

TikTok launches $2 billion Creator Fund to support content creators on the app
(July 2020)

TikTok sues the Trump admin. over "unconstitutional" ban attempt

Trump passes an executive order to halt transactions between US & ByteDance
(Aug. 2020)

TikTok says it is working to combat hate speech
(Oct. 2020)

FTC issues order to TikTok requesting data collection and use practices
(Dec. 2020)

TikTok settles $92 million class-action lawsuit
President Biden pauses Trump's TikTok ban
(Feb. 2021)

TikTok sued by former children's commissioner for England over misuse of child data
(Apr. 2021)

🟥 TikTok trouble

🟦 TikTok progress

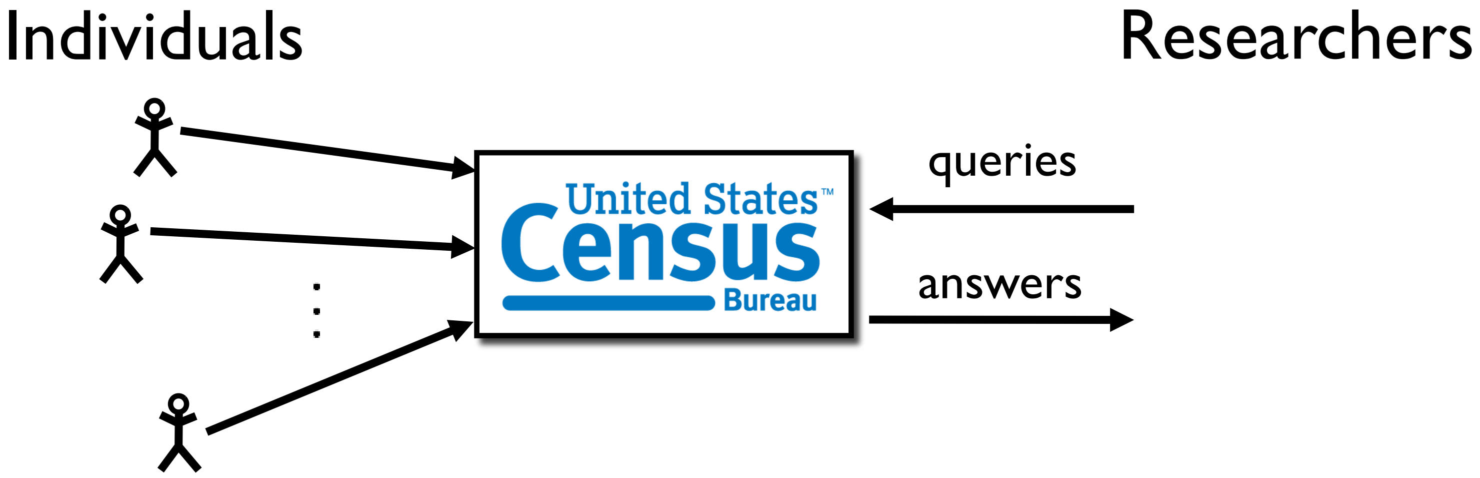
Chan, N. K., Su, C. C., & Shore, A. (2023). Shifting platform values in community guidelines: Examining the evolution of TikTok's governance frameworks. *New Media & Society.*

Pinning Down "Privacy"

Adam Smith

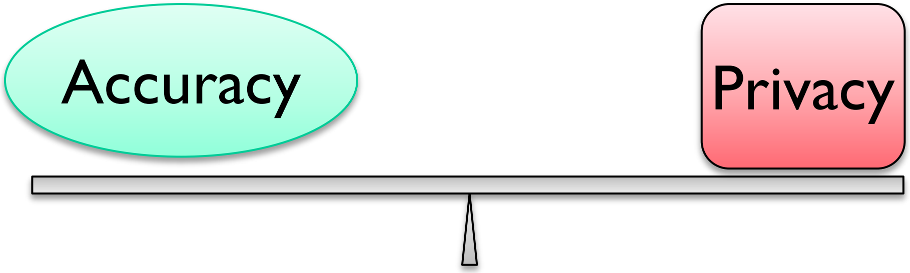
Professor
Computer Science, CAS
Electrical and Computer Engineering, CoE
Faculty of Computing and Data Sciences

Privacy in Statistical Databases



Large collections of personal information

- census data
- medical/public health
- social networks
- Education



My research: Rigorous foundations and analysis

First attempt: Remove obvious identifiers

"AI recognizes blurred faces"
[McPherson Shokri Shmatikov '16]

[Gymrek McGuire Golan Halperin Erlich '13]

- Everything is an identifier
- Potential attackers have other information
- "Anonymization" schemes are regularly broken

[Pandurangan '14]

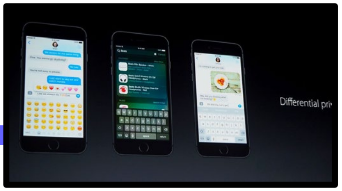
On Taxis and Rainbows

Lessons from NYC's improperly anonymized taxi logs

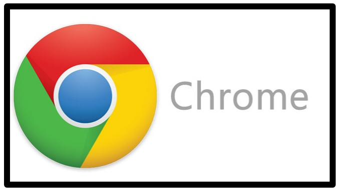
Images: whitehouse.gov, genesandhealth.org, medium.com

Rigorous Approaches to Privacy

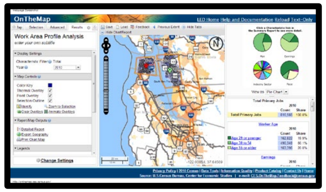
- Definitions
  - ➢ Pinning down “privacy”
- Algorithms: what can we compute privately?
  - ➢ Fundamental techniques
  - ➢ Specific applications
- Attacks: “Cryptanalysis” for data privacy
  - ➢ Impossibility results
- Implications for other areas



Apple



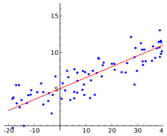
Google



US Census

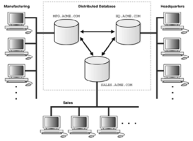
Algorithms

Crypto, security



Statistics, learning

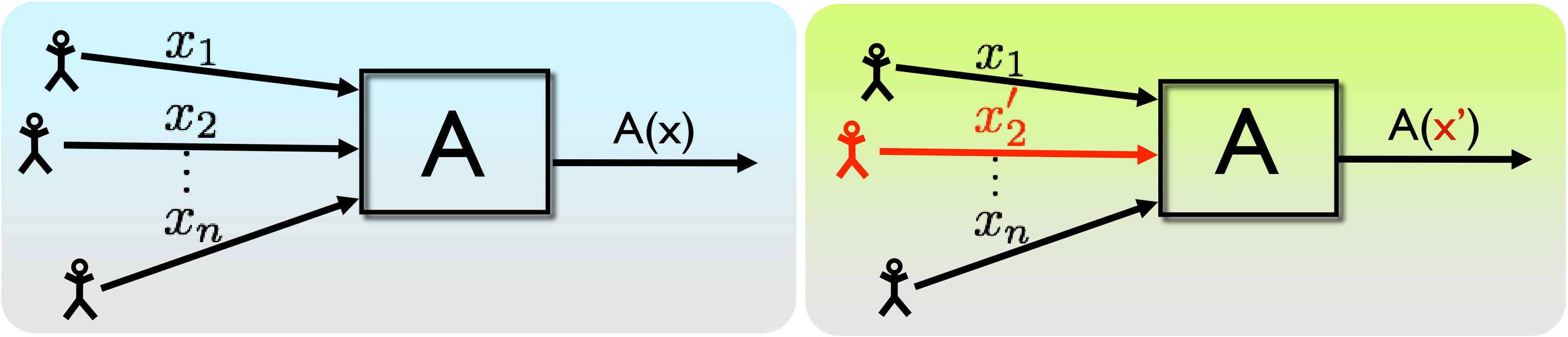
Game theory, economics



Programming languages

Law, policy

Differential Privacy *[Dwork, McSherry, Nissim, S.]*



- A thought experiment
  - Change one person's data (or remove them)
  - Will the probabilities of various outputs change much?
- Differential privacy implies:

  No matter what you know ahead of time,

  > You learn (almost) the same things about me
  > whether or not my data are used

The Social Efficiency of Fairness

Marshall Van Alstyne

Allen & Kelli Questrom Professor of IS
Questrom School of Business

GenAI

Sara Silverman
John Grisham
Authors’ Guild
Writer’s Guild

vs

Stability AI
Meta
Midjourney
Open AI

- *© Infringement*
- *Unjust enrichment*
- *Unfair competition*

NY Times
Reddit
StackOverflow
Twitter …

All close their APIs

Problem

Defendants Win

✗ Destroy Livelihoods

✗ Stop Sharing, Close Access

Plaintiffs Win

✗ Shut down tech innovation

✗ How negotiate with 10M © owners?!?

Solution

Permissionless Innovation + Fair Compensation
(*Shapley Formula : Nobel 2012*)

$$\phi_i(v) = \frac{1}{|N|} \sum_{S \subseteq N \setminus \{i\}} \binom{|N|-1}{|S|}^{-1} (v(S \cup \{i\}) - v(S))$$

You get your share of value for *any* set of assets to which you add value

Formal Proof

1. Fairness increases the rate of innovation.
   - Sources are more willing to share data.
   - Welfare improves both in the absolute sense of enabling new projects and in the relative sense of reordering the social sort order of which projects agents undertake.

2. Fairness (liability) rather than property rules can be more conducive to innovation based on information reuse and recombination.

- "***Social Efficiency of Fairness***" on ssrn.com
- Working with Industrial Firms to build the markets

THANK YOU!

UPCOMING EVENTS

Learn more & RSVP: bu.edu/research/events
Topic ideas & feedback: bu.edu/research/topic-ideas

RESEARCH ON TAP

Measuring Corporate Impacts on the Environment & Society
November 13, 2023 | 4-6 pm

Health Data Science
November 29, 2023 | 4-6 pm

RESEARCH HOW-TO

What the X? How to Make the Most of Social Media to Promote your Research in 2023
October 26, 2023 | 12-1:30 pm

Meet the Alzheimer's Association
November 8, 2023 | 12-1 pm