DSC 291 Privacy-sensitive Data Systems (week 1b)

Haojian Jin

Logistics

1. Canvas access
2. Enrollment
3. Discussion leader signup
4. Find your partners (2-3 people each team).
5. Arrange a meeting with me in the next two weeks.

Due this Sunday!

Recap

- Week 2: Apple AirTag, Contact Tracing
- Week 3: Android permission, Browser cookie consent
- Week 4: P3P
- Week 5: Target pregnancy prediction
- Week 6: US Census Bureau
- Week 7: Facebook cambridge analytic
- Week 8: Big brother privacy
- Week 9: Dark patterns

Today's topic: Why is privacy hard?

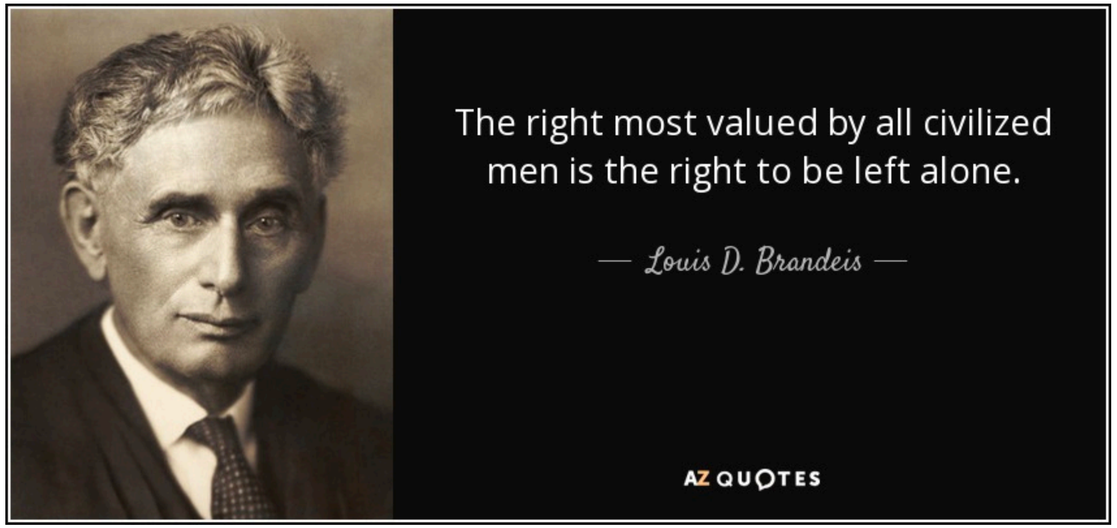
- 10 reasons [1]
  - Lots of examples from smartphones and web
  - Lessons learned from community's previous failures
  - Directions for research and practice

[1] Adapted from https://cacm.acm.org/blogs/blog-cacm/235401-why-is-privacy-so-hard/fulltext

#1 Privacy is a broad and fuzz term

- Privacy is a broad umbrella term that captures concerns about our relationships with others.
  - The right to be left alone
  - Control and feedback over one's data
  - Anonymity (most popular among researchers)
  - Presentation of self (impression management)
  - Right to be forgotten
  - Contextual integrity (take social norms into account)
- Each definition leads to a different way of handling privacy.

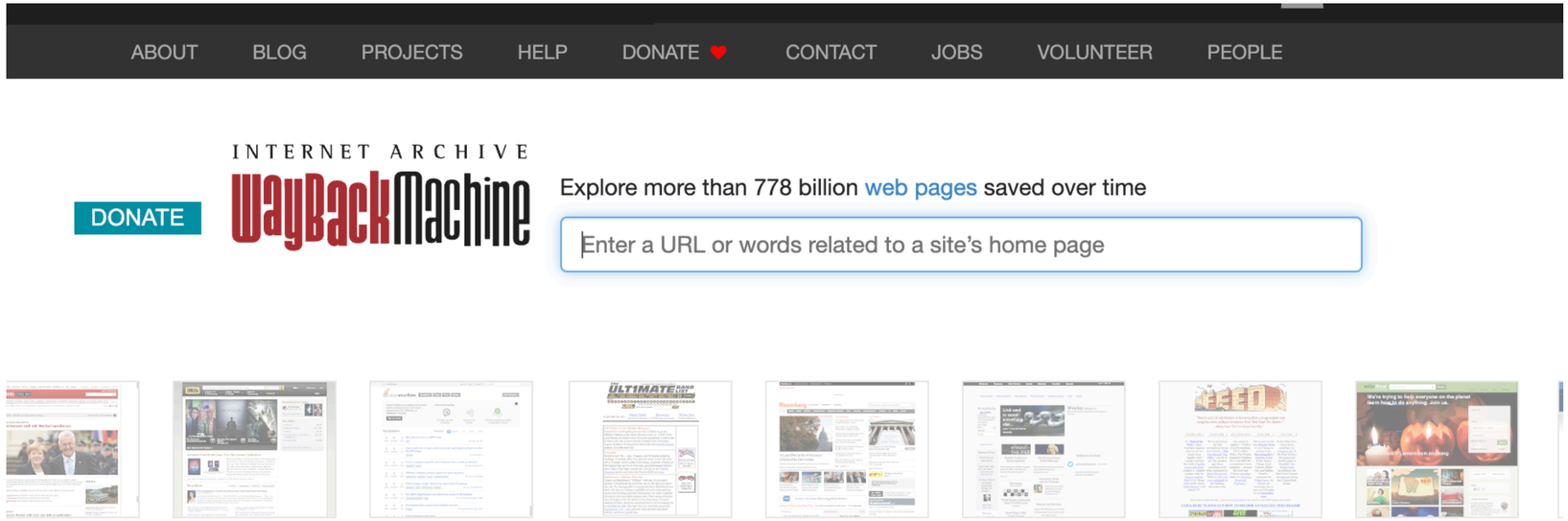
Right to be left alone: Do not call list, blocking.



Why Right to be forgotten?

https://www.cnn.com/2022/05/20/us/ryan-last-suicide-sextortion-california/index.html

Wayback machine



Demo link: https://web.archive.org/web/201311010000000*/shift-3.com

Right to be forgotten: delete from search engine, services

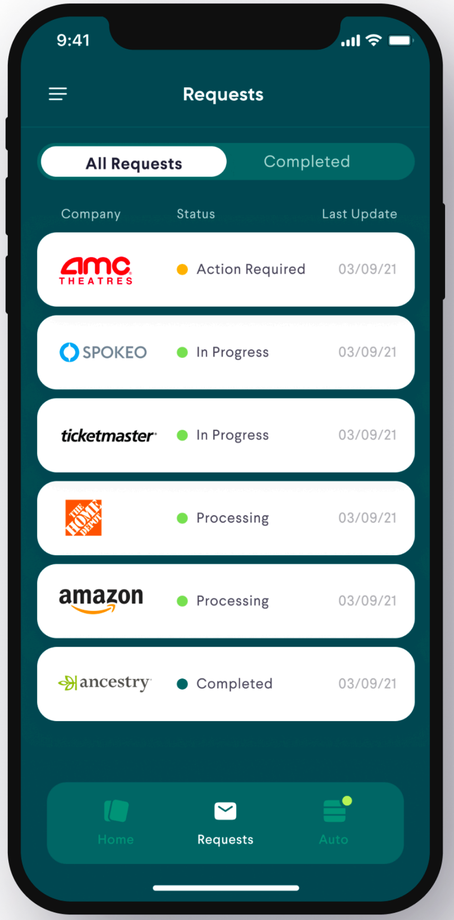
Consumers' "Right to Delete" under US State Privacy Laws

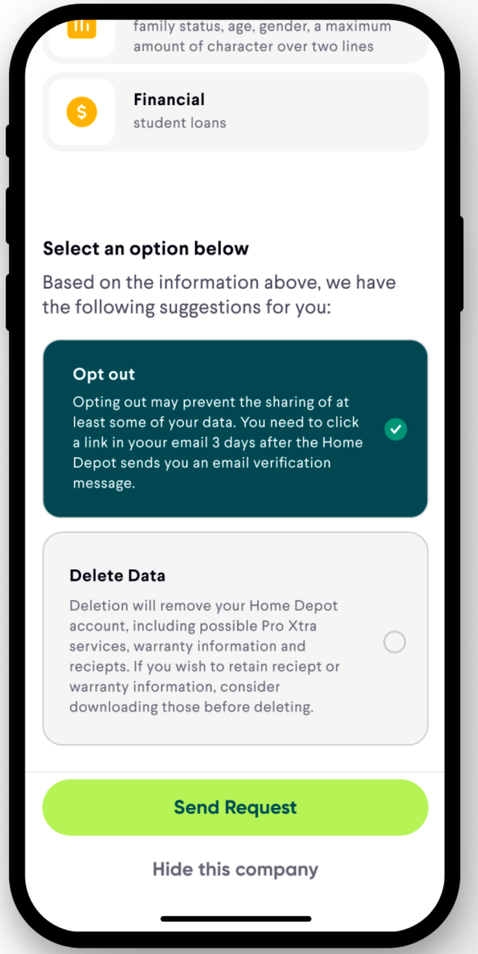
By Glenn A. Brown on March 3, 2021

Posted in California Privacy Rights Act (CPRA), CCPA, US

Businesses have to ***confirm receipt of a deletion request*** within 10 business days … ***permanently and completely erasing the personal information*** within 45 calendar days.

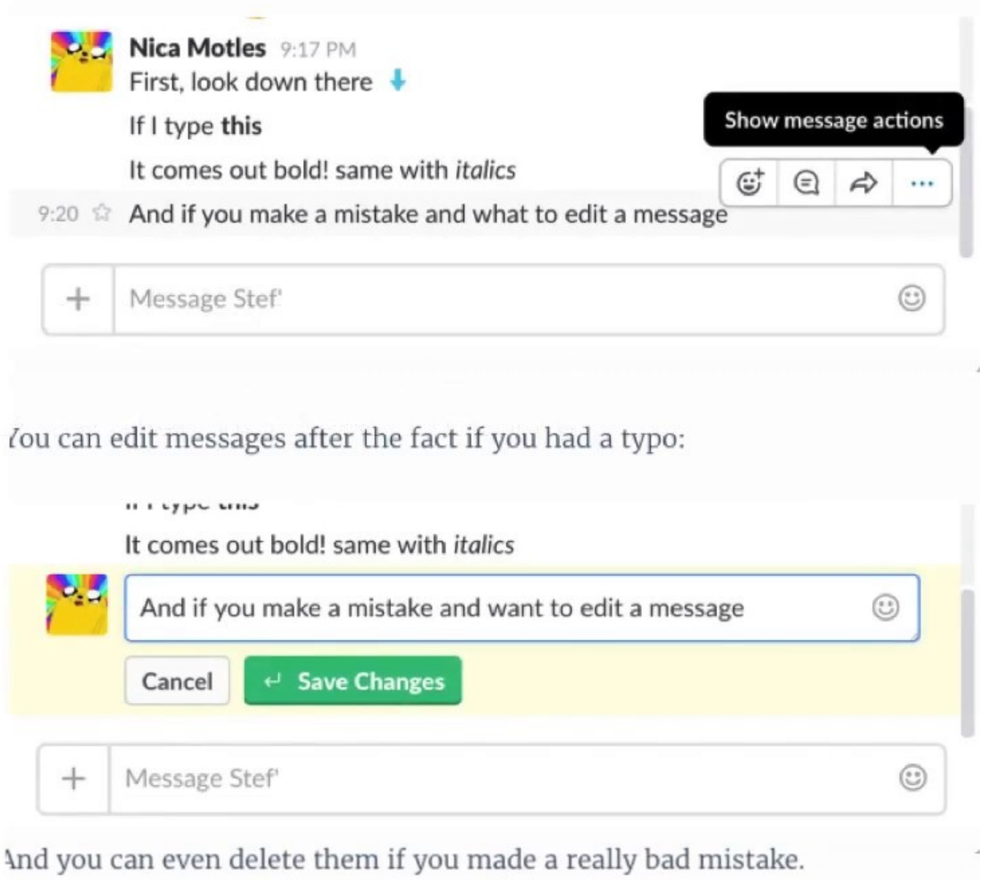
Permission slip by consumer reports





Demo: Enhancing Email Functionality using Late Bound Content

Self destruct - SnapChat

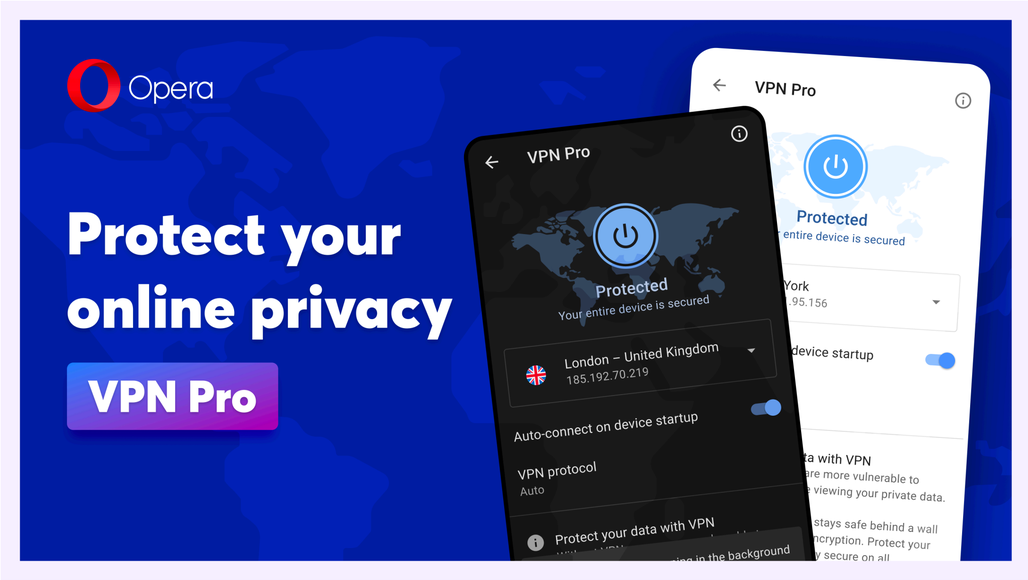


Continuous edit - Slack

Many more questions!

- Deleting data is hard.
  - Data transmission and storage are so hard.
  - Distributed system, multiple copies.
  - Data stored across different countries.
  - Different user preferences.
  - ……

Anonymity: VPN



Personally Identifiable Information (PII)

- Data privacy is primarily about how orgs collect, use, and protect sensitive data (PII)
  - Ex. Name, street address, unique IDs, pictures
    - Vague definitions.
  - Rules about data use, privacy notices, developer understanding.

Fair Information Practices

- Notice / Awareness
- Choice / Consent
- Access / Participation
- Integrity / Security
- Enforcement / Redress

A CONSUMER INTERNET PRIVACY

BILL of RIGHTS

The Obama Administration believes America must apply our timeless privacy values to the new technologies and circumstances of our times. Citizens are entitled to have their personal data handled according to these principles.

Individual Control
Consumers have a right to exercise control over what personal data companies collect from them and how they use it.

Access and Accuracy
Consumers have a right to access and correct personal data in usable formats, in a manner that is appropriate to the sensitivity and risk associated with the data.

Transparency
Consumers have a right to easily understandable and accessible information about privacy and security practices.

Focused Collection
Consumers have a right to reasonable limits on the personal data that companies collect and retain.

Respect for Context
Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent.

Accountability
Companies should be accountable to enforcement authorities and consumers for adhering to these principles.

Security
Consumers have a right to secure and responsible handling of personal data.

LEARN MORE AT WHITEHOUSE.GOV

An AI Bill of Rights

Safe and Effective Systems

Algorithmic Discrimination Protections

Data Privacy

Notice and Explanation

Human Alternatives, Consideration, and Fallback

Applying the Blueprint for an AI Bill of Rights

Download the Blueprint for an AI Bill of Rights ↗

https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Procedural + Consequential Perspectives

- Procedural perspective
  - Did you follow this set of rules? •
  - Did you check off all of the boxes? •
  - Somewhat hard to measure too (Better? Worse?) •
- Outcome-oriented
  - Computer scientists love this approach, can measure!
  - K-anonymity, differential privacy
    - However, these approaches tend to be narrow

#2 Technological Capabilities Rapidly Growing

- Data gathering easy and pervasive
  - Everything one the web (web clicks, browsing history, video watching)
  - Sensors (cameras, microphones, standby)
  - Transaction data (online shopping)

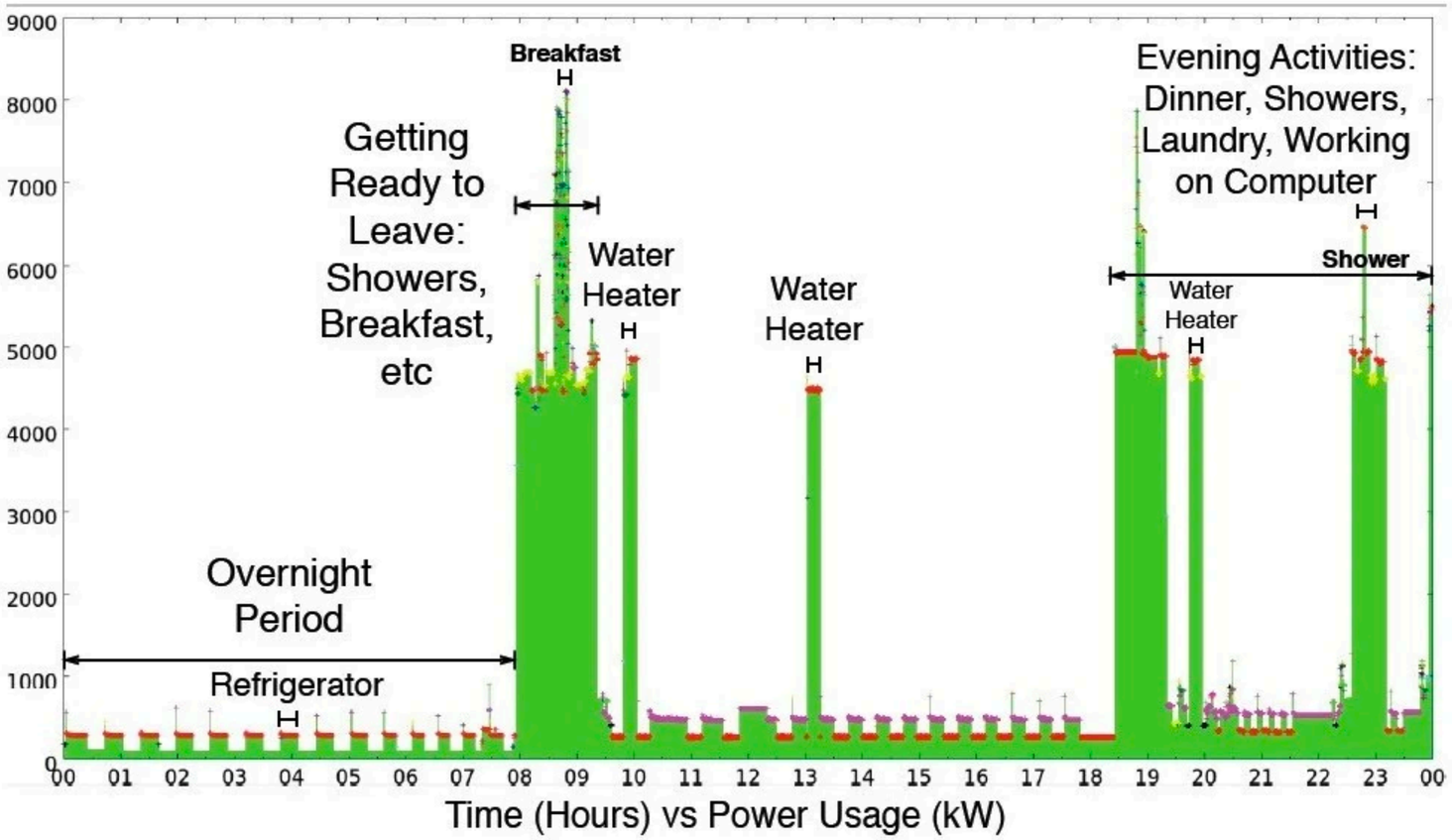
#2 Technological Capabilities Rapidly Growing

- Data storage and querying bigger and faster

1TB hard drive in 1937.
the largest vertical letter file in the world.
4000 SqFt.
with over 3000 drawers
10 feet long managed by 20 workers.
Access speed was ~3 minutes per KB.

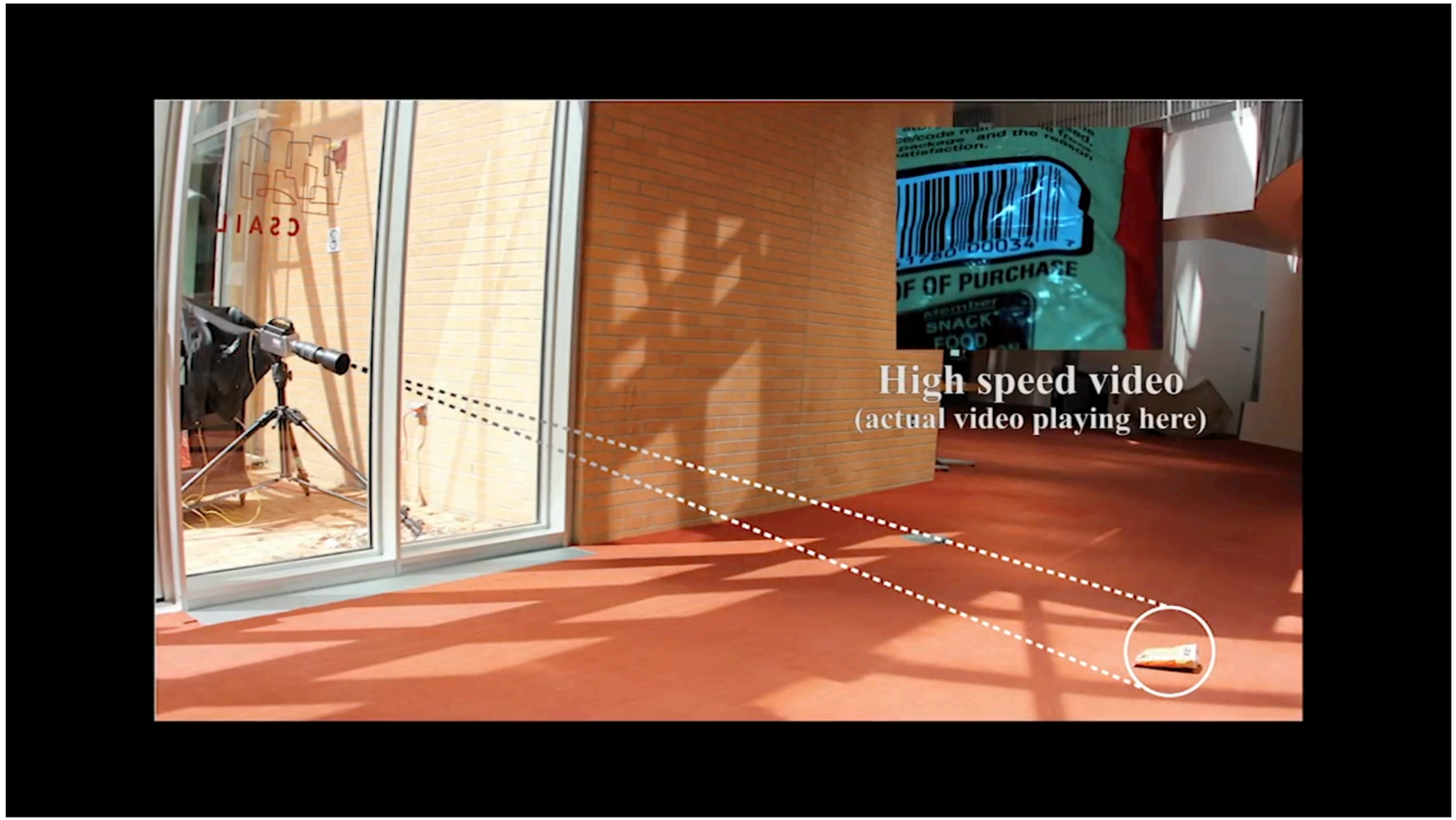
#2 Technological Capabilities Rapidly Growing

- Inferences are becoming more powerful.



Private Memoirs of a Smart Meter, BuildSys 2010

Sky is the limit.



When sounds hits an object, it causes that object to vibrate.

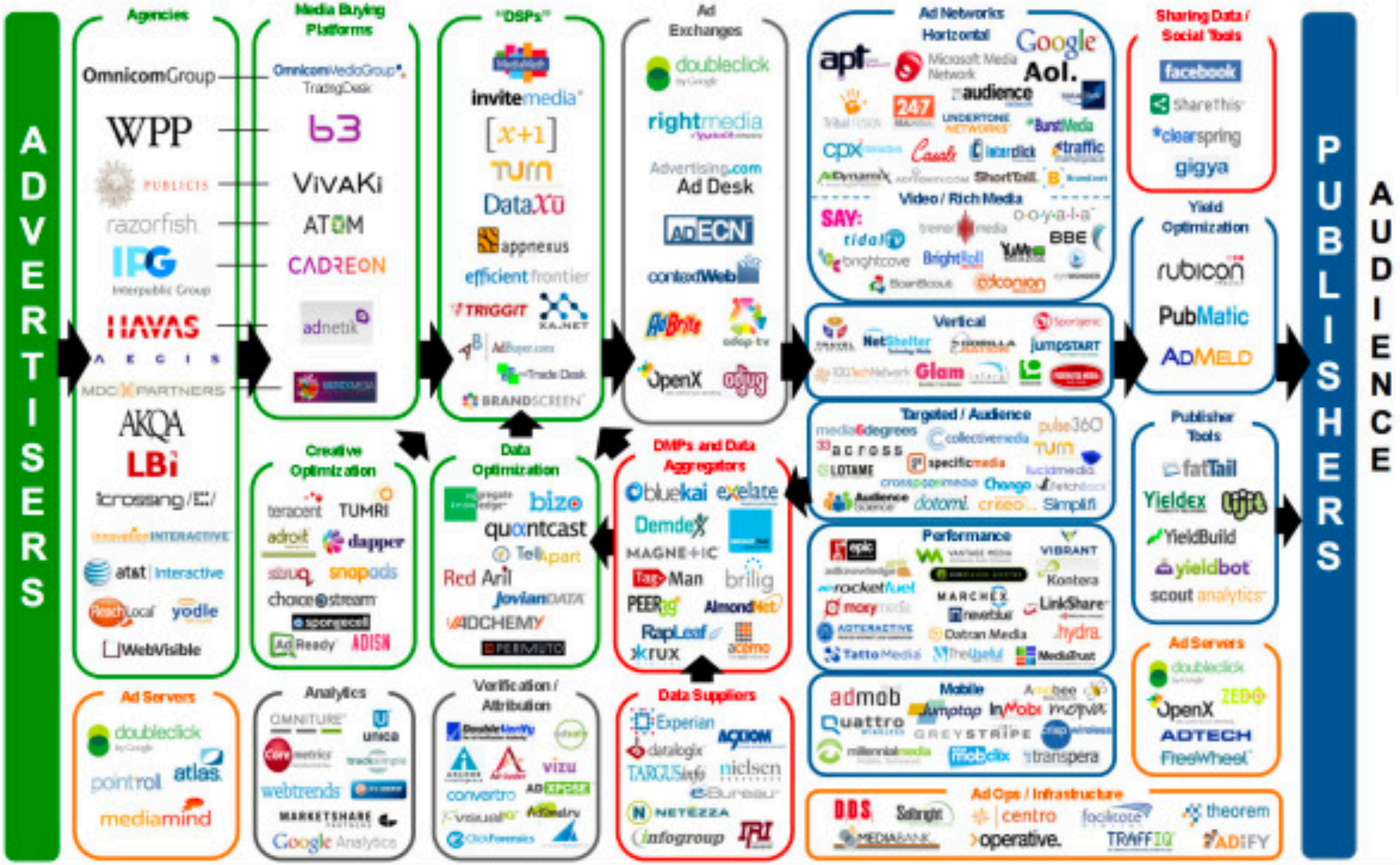
#2 Technological Capabilities Rapidly Growing

- More inferences
  - Are you healthy (Microphone/Camera)?
  - Are you pregnant (Purchases at Target)?
  - Personality type (Instagram, smartphone).
  - Are you depressed or not? (Smartphone)
  - Workplace social network + performance (Call logs)
  - In-home activities (Laptop usages)
  - …

#2 Technological Capabilities Rapidly Growing

- Data sharing more widespread
  - Lots of companies collecting and sharing data about you.
  - It is hard to explain to end users.
  - It is hard to track them.

Display Advertising Technology Landscape

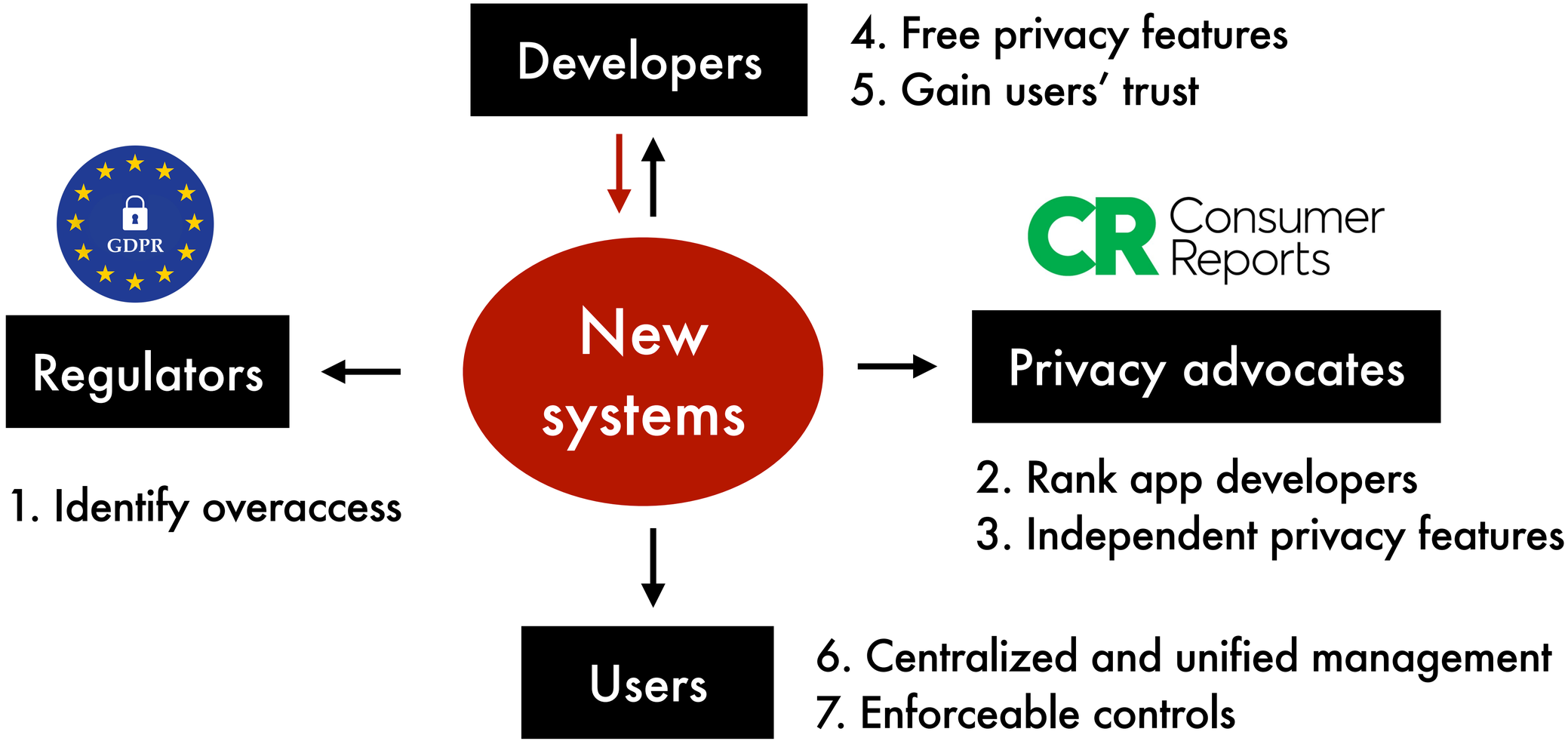


tkawaja@lumapartners.com

#3 Strong Incentives for Companies to Collect Data

- More data -> better ML models -> bottom line
  - Increasing relevance of online ads worth millions
  - “Post-purchase monetization”

Let the good privacy drive out the bad privacy

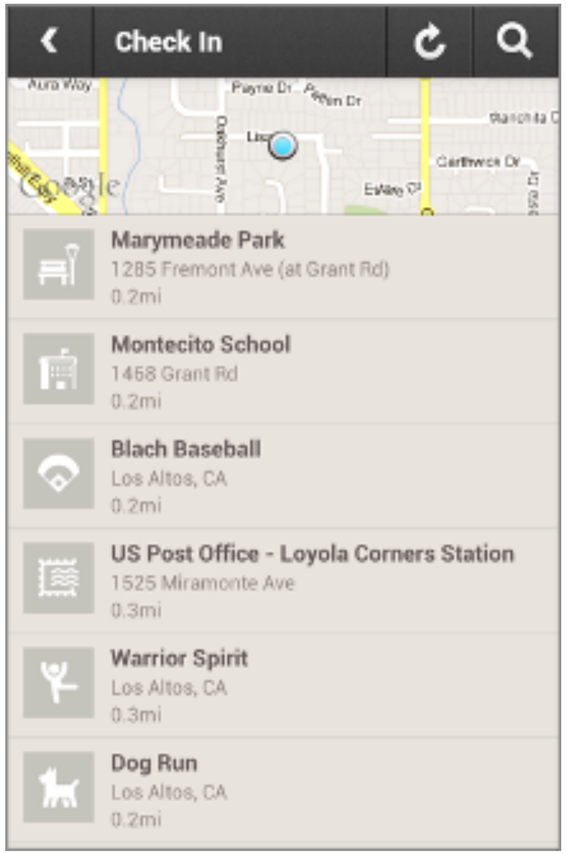


#4A Same Device, Different Perspectives

- Locator badges for nurses
- Hospital administration's perspective
  - Coordination ("Where is Alice?")
  - Protect from spurious claims ("Nurse never came")
- Nurses' perspective
  - Surveillance ("Tracking how long I was in restroom")
- Clear value + management trusted -> ok •
- Existing tensions -> rejected badges

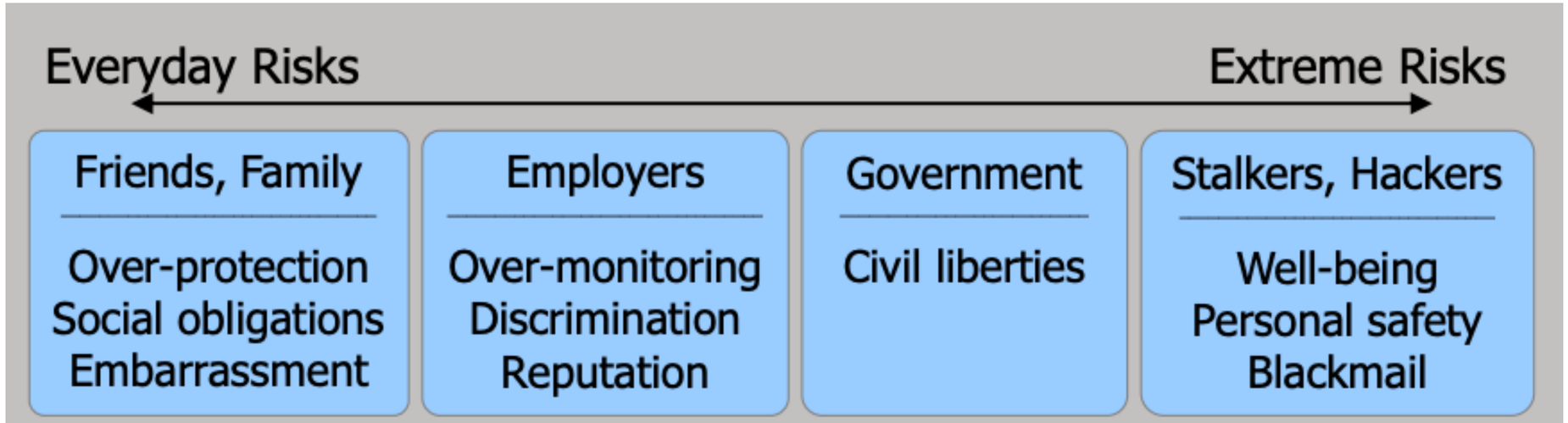
#4B Same Data, Different Perspectives

- Data may be ok in one context, creepy in another



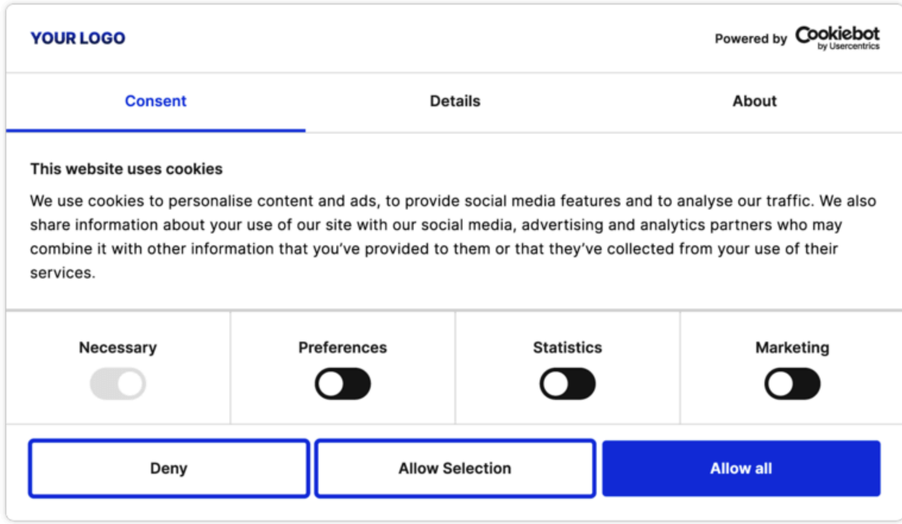
#5 Wide Range of Privacy Risks

- It's not just Big Brother or just corporations •
- Privacy is about our relationships with every other individual and organization out there
  - Will need different solutions for different relations
  - Ex. Data privacy useless for friends + family

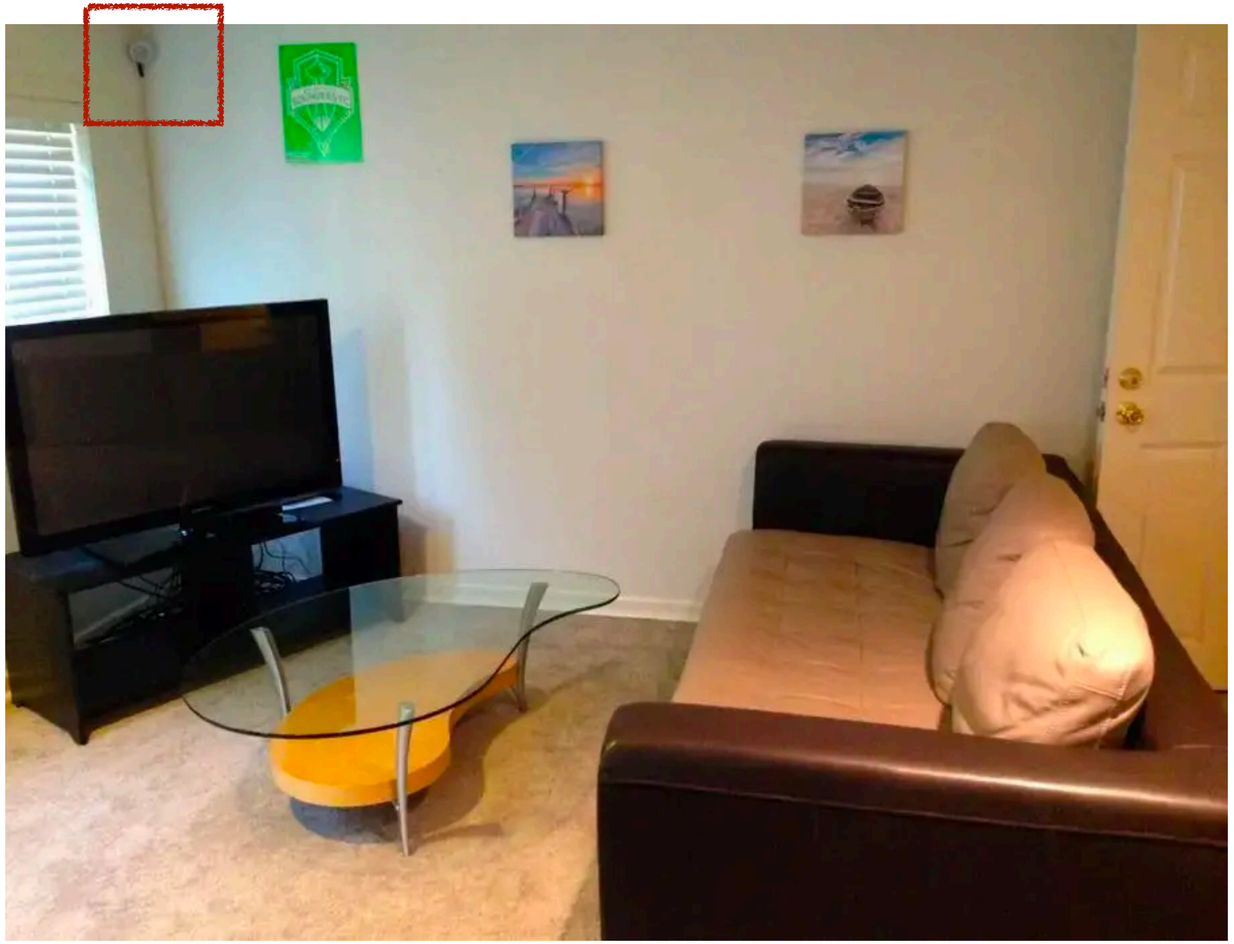


#6 Burden on End-Users is Too High

- Individuals have to make too many decisions
  - Is this device good for privacy?
  - Should I install this app?
  - What are all the settings I need to know?
  - What are all the terms and conditions?
  - Trackers, cookies, VPNs, anonymizers, etc



Cameras in Airbnb

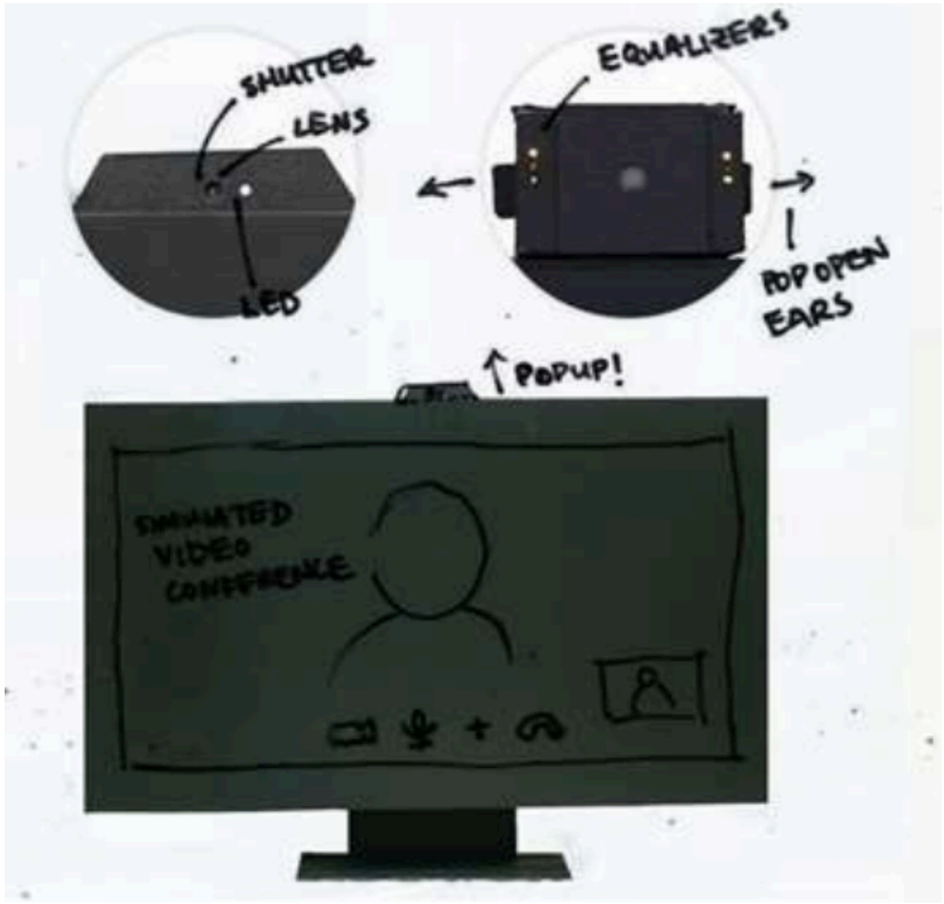


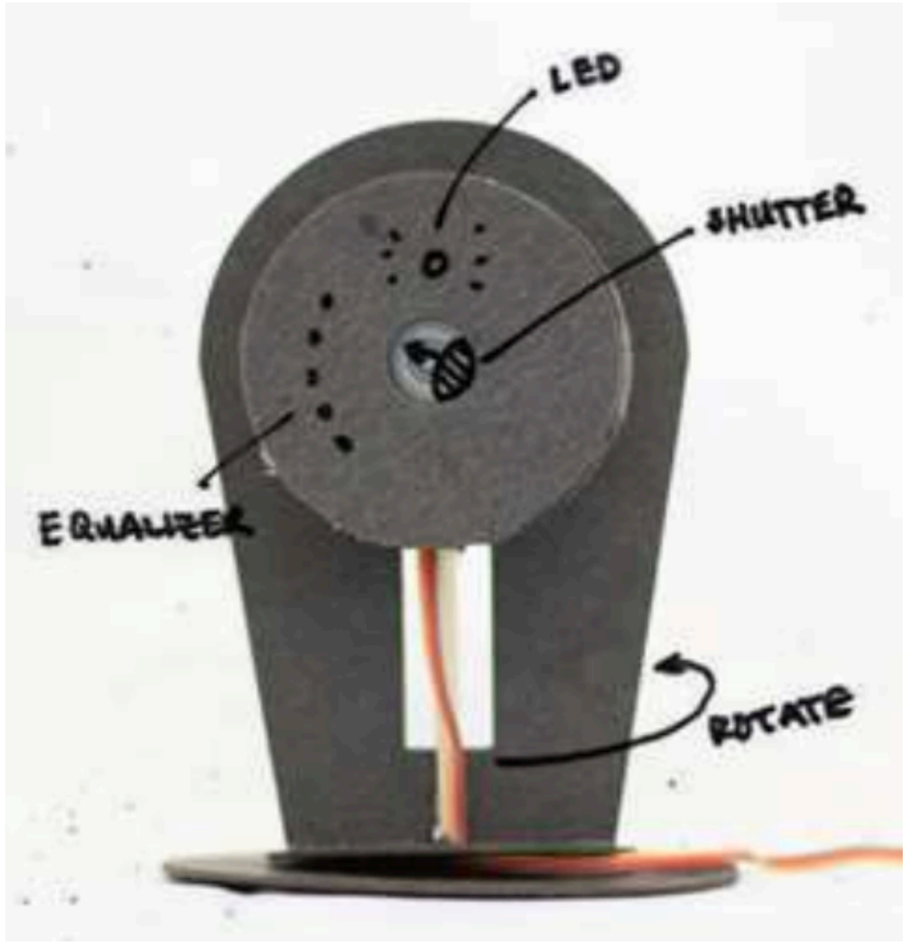
Travel

AIR SPY Airbnb guest finds surveillance camera inside his rented apartment but is told he ‘consented’ to it as it appeared in photos of the property

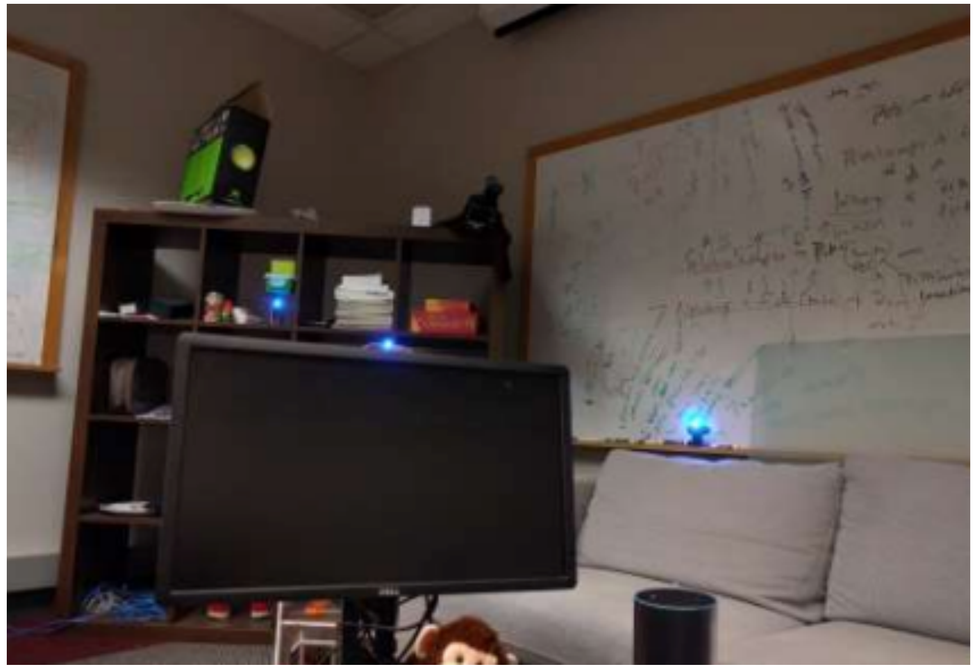
He claimed he was not made aware of the internal camera while booking

Design patterns

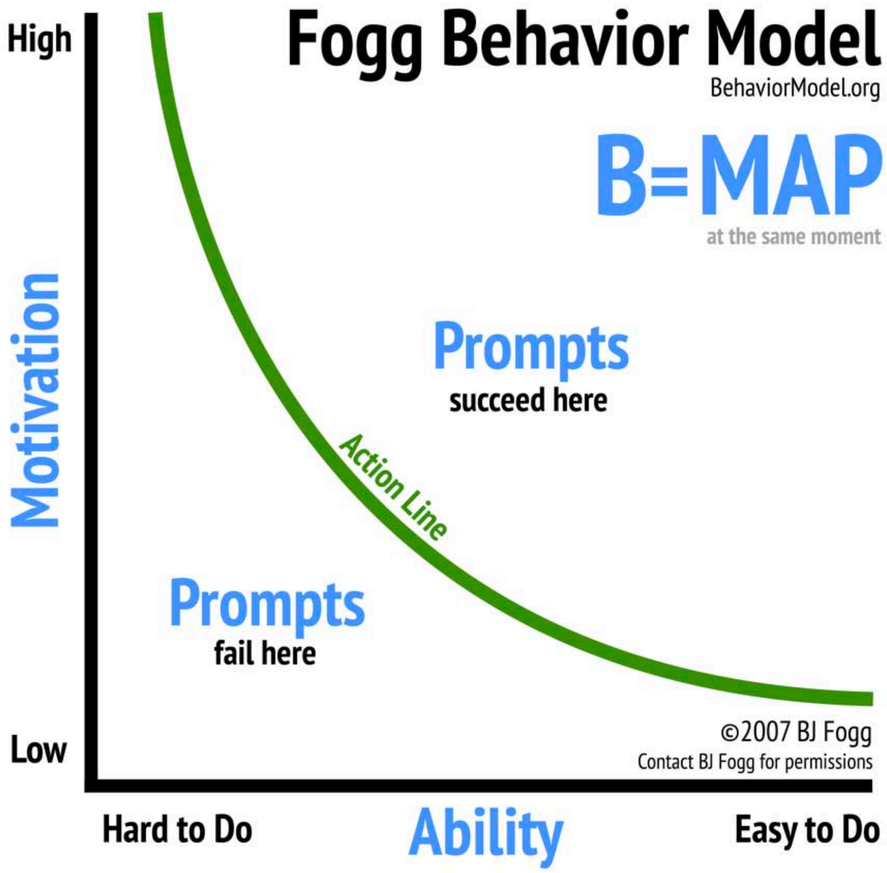




Can we make it easy to locate devices?



#7 Low Knowledge, Awareness, Motivation by Practitioners



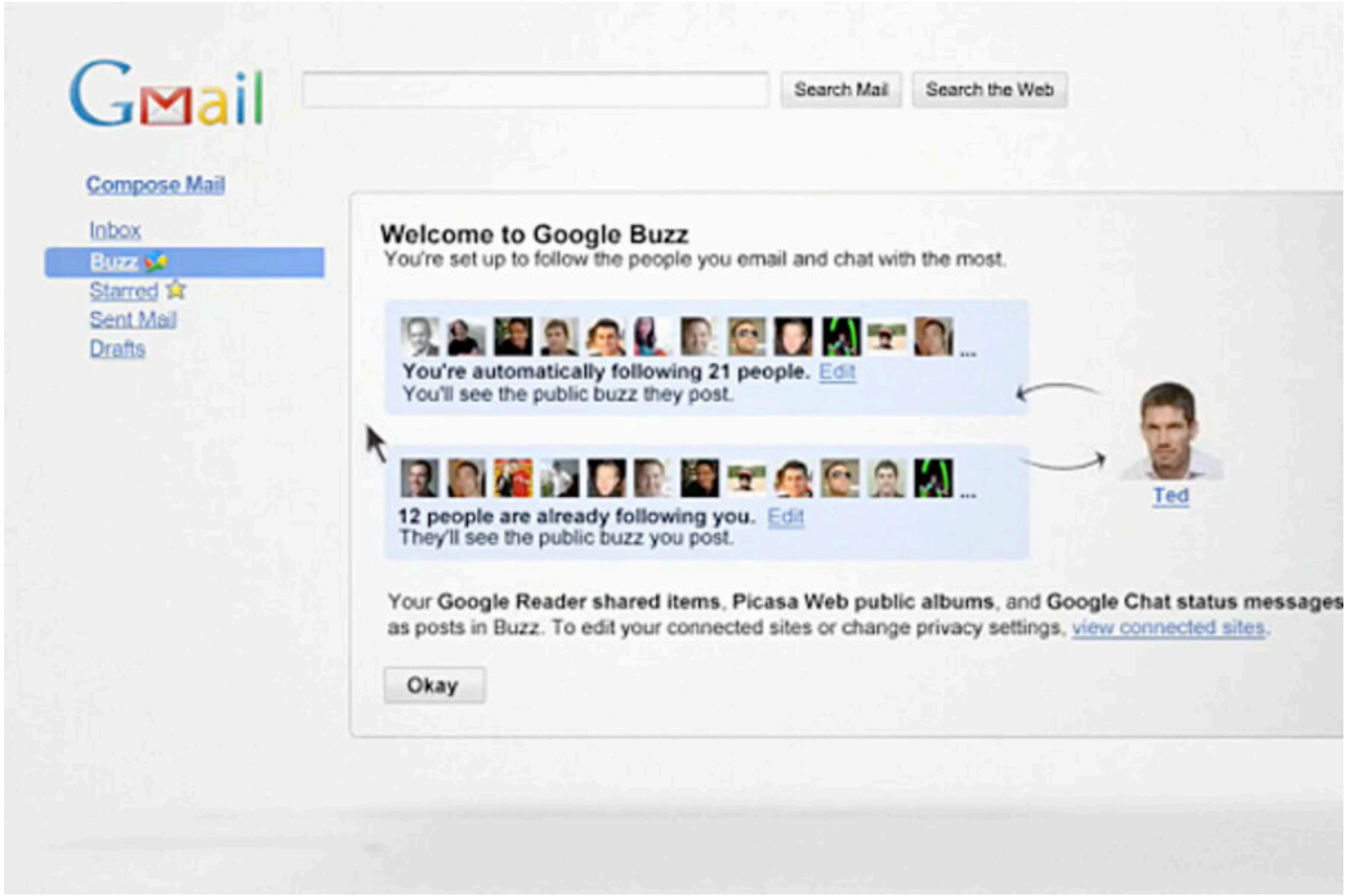
https://behaviormodel.org/

#7 Low Awareness

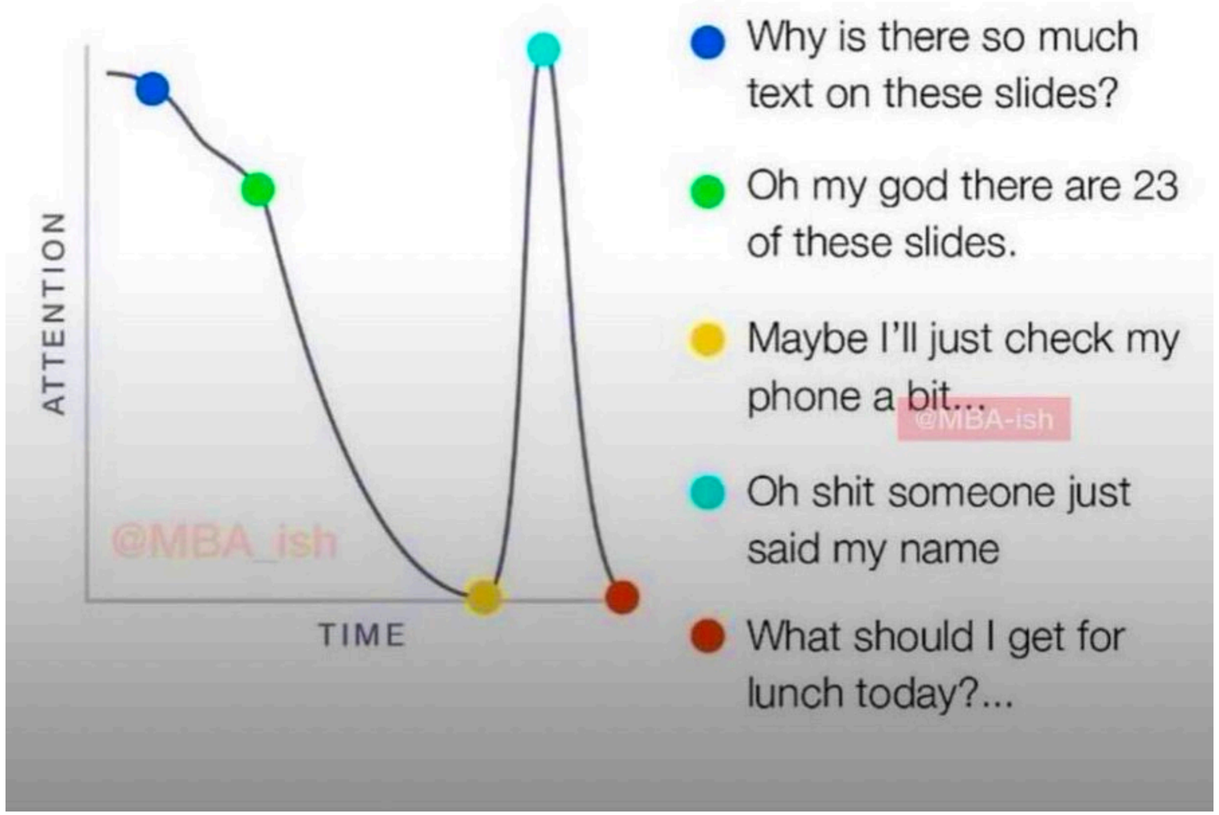
- Low awareness of privacy issues in their apps.
  - Google Buzz
  - Zoom Attention
  - Google Meeting
  - OKCupid
  - PrivacyGrade

https://behaviormodel.org/

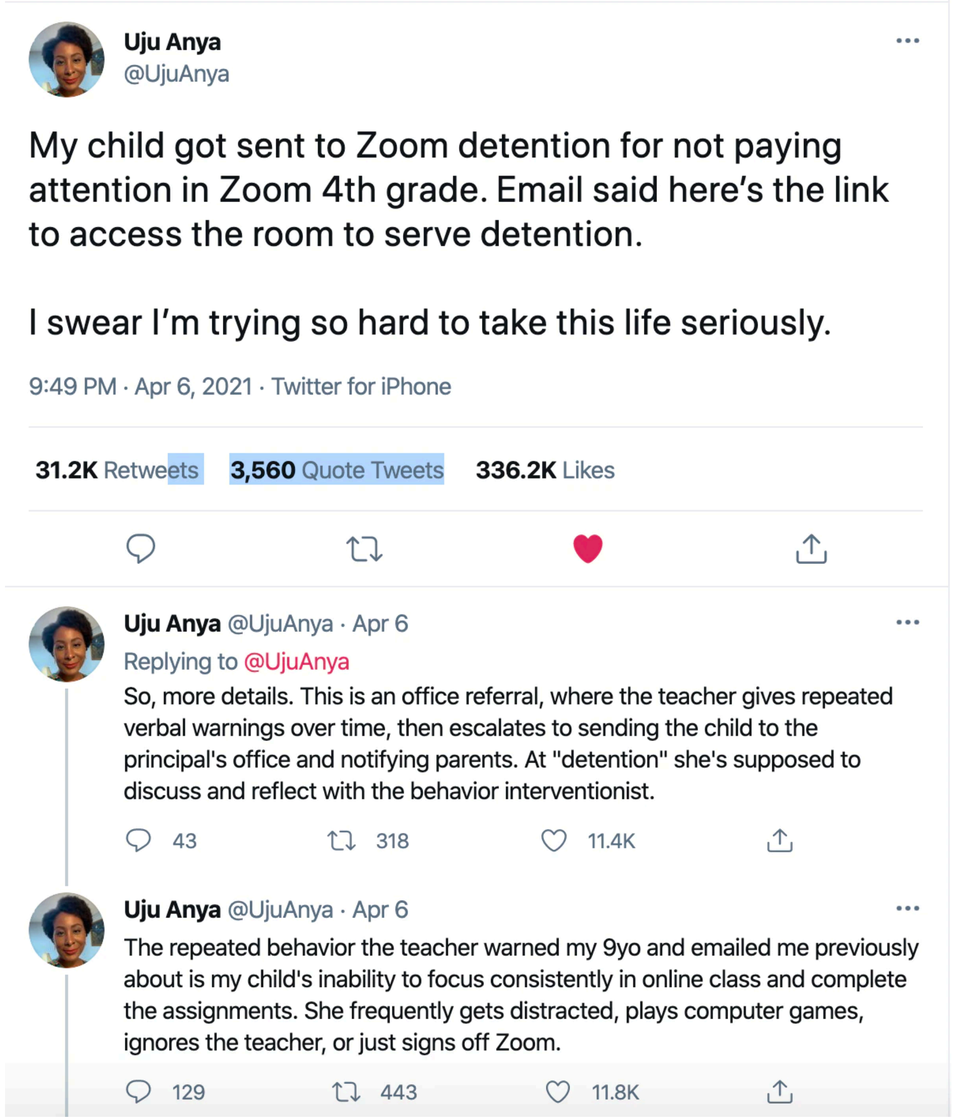
Google Buzz


Gmail
Search Mail
Search the Web
Compose Mail
Inbox
Buzz
Starred
Sent Mail
Drafts
Welcome to Google Buzz
You're set up to follow the people you email and chat with the most.
You're automatically following 21 people. Edit
You'll see the public buzz they post.
12 people are already following you. Edit
They'll see the public buzz you post.
Ted
Your Google Reader shared items, Picasa Web public albums, and Google Chat status messages as posts in Buzz. To edit your connected sites or change privacy settings, view connected sites.
Okay

Zoom attention



Zoom attention



Uju Anya
@UjuAnya

My child got sent to Zoom detention for not paying attention in Zoom 4th grade. Email said here's the link to access the room to serve detention.

I swear I'm trying so hard to take this life seriously.

9:49 PM · Apr 6, 2021 · Twitter for iPhone

31.2K Retweets **3,560** Quote Tweets **336.2K** Likes

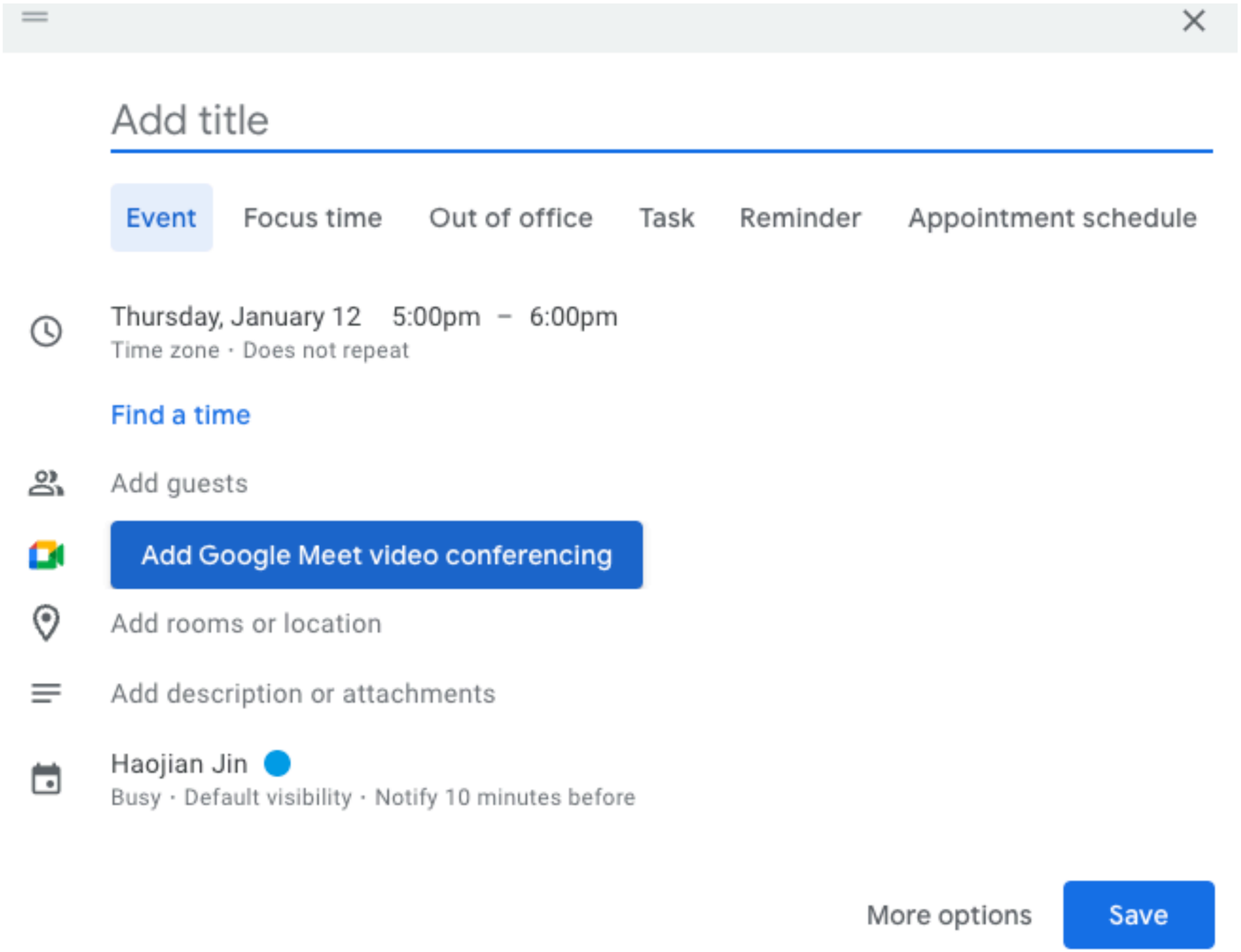
Uju Anya @UjuAnya · Apr 6
Replying to @UjuAnya
So, more details. This is an office referral, where the teacher gives repeated verbal warnings over time, then escalates to sending the child to the principal's office and notifying parents. At "detention" she's supposed to discuss and reflect with the behavior interventionist.

43 318 11.4K

Uju Anya @UjuAnya · Apr 6
The repeated behavior the teacher warned my 9yo and emailed me previously about is my child's inability to focus consistently in online class and complete the assignments. She frequently gets distracted, plays computer games, ignores the teacher, or just signs off Zoom.

129 443 11.8K

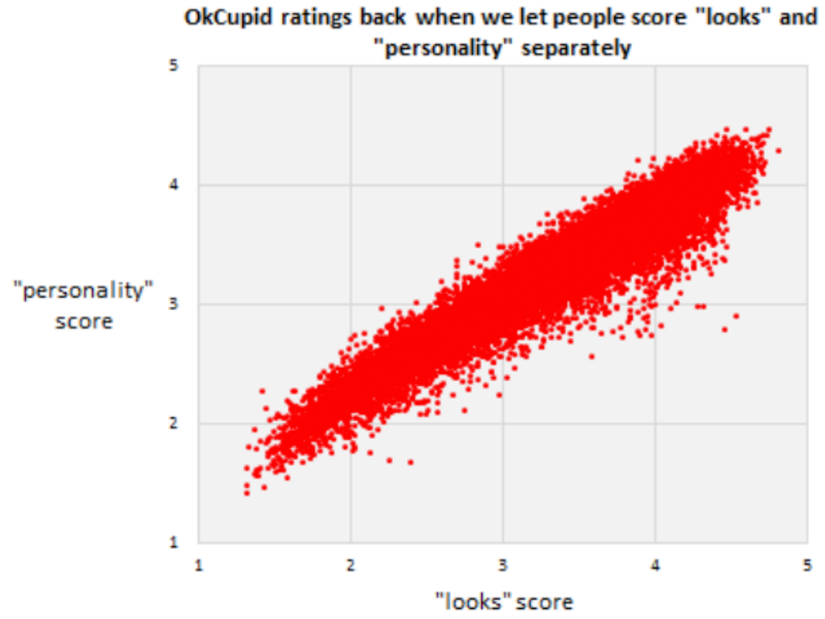
Google Meet

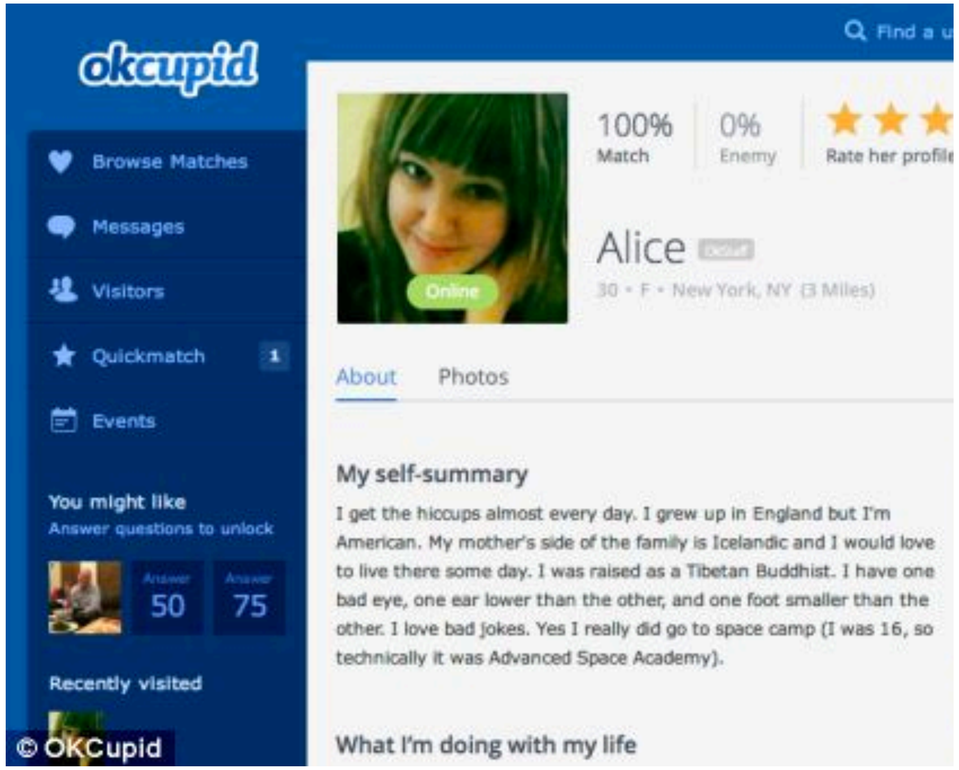

Add title
Event
Focus time
Out of office
Task
Reminder
Appointment schedule
Thursday, January 12 5:00pm – 6:00pm
Time zone · Does not repeat
Find a time
Add guests
Add Google Meet video conferencing
Add rooms or location
Add description or attachments
Haojian Jin
Busy · Default visibility · Notify 10 minutes before
More options
Save

OKCupid | Facebook Social Contagion

OKCupid: we experiment on users. Everyone does

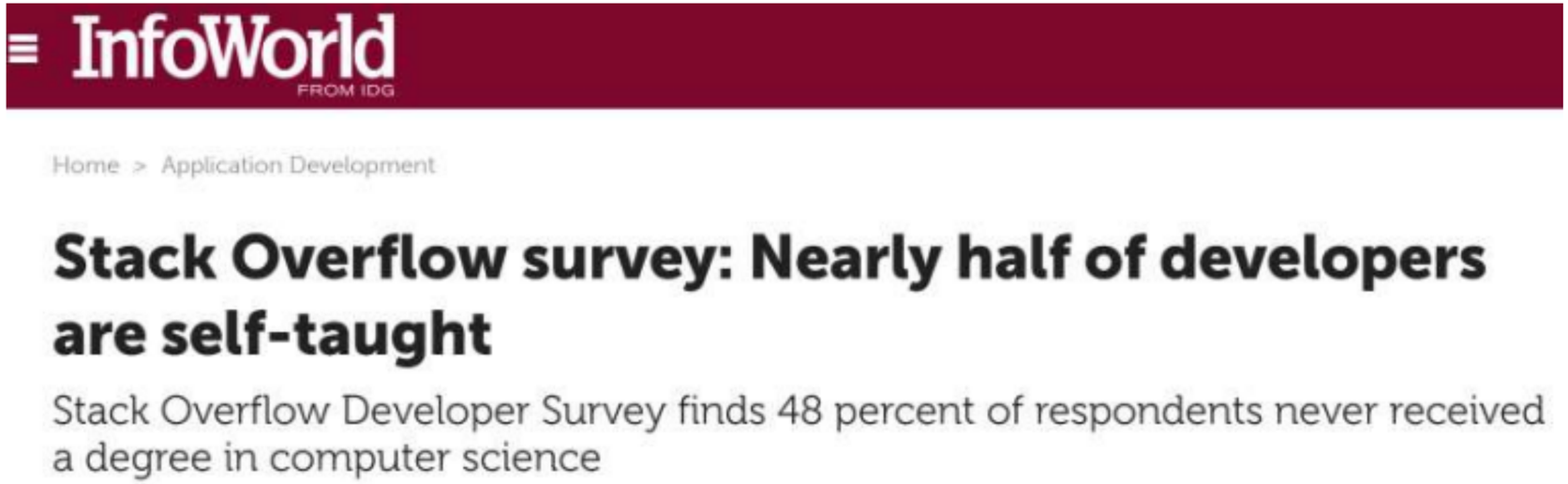
Dating service's co-founder claims it lied to users to test theory about compatibility



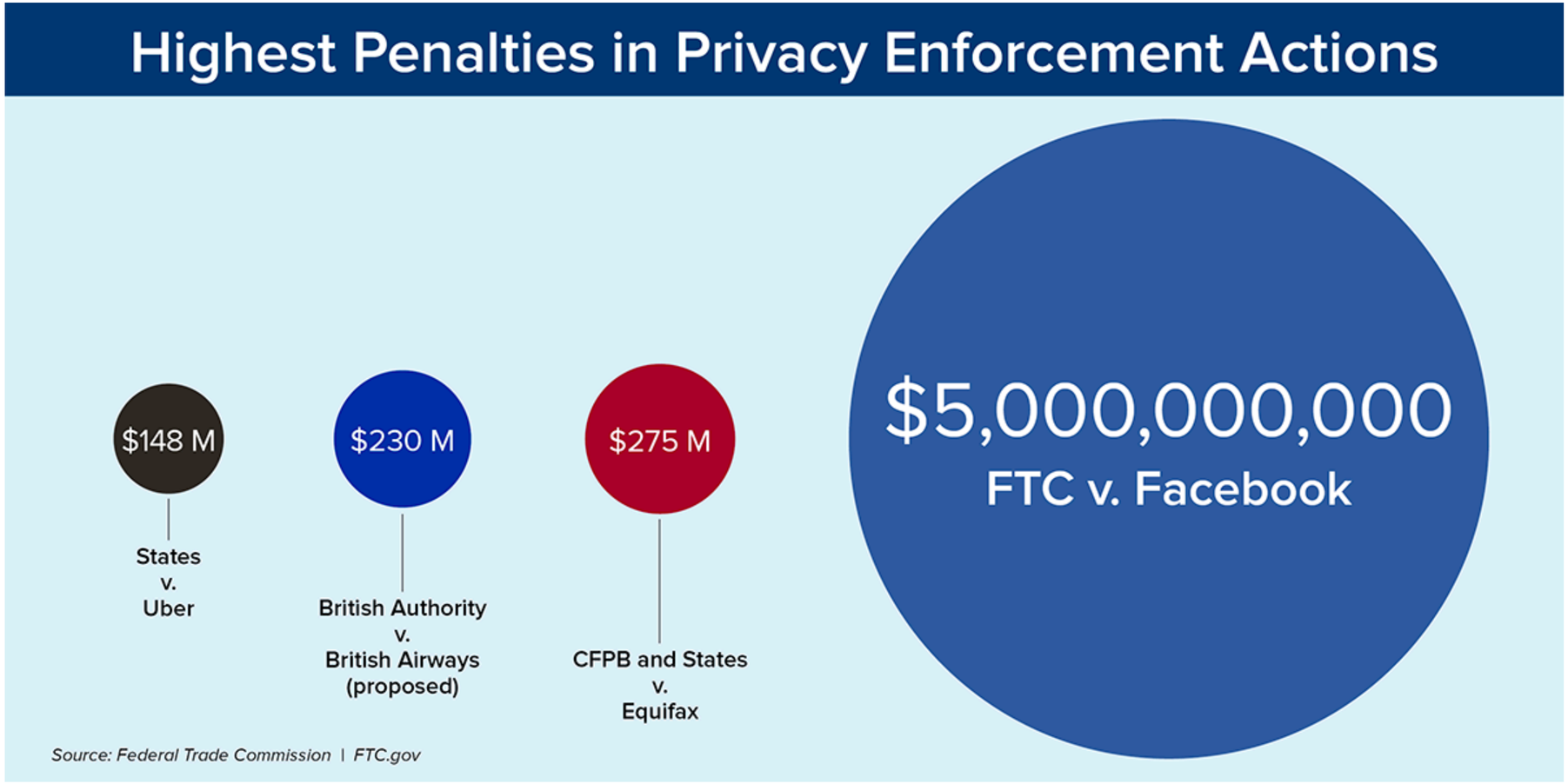


#7 Low Knowledge

- Developers have low knowledge of privacy.
  - What is PII?
  - How should I interpret policies?



Low Motivation (1)



https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook

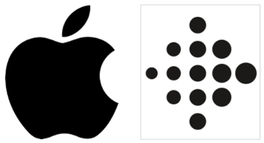
Low Motivation (2)

Social

Shopping

Communication

Health & Fitness



Entertainment

Travel

Search

Home & House

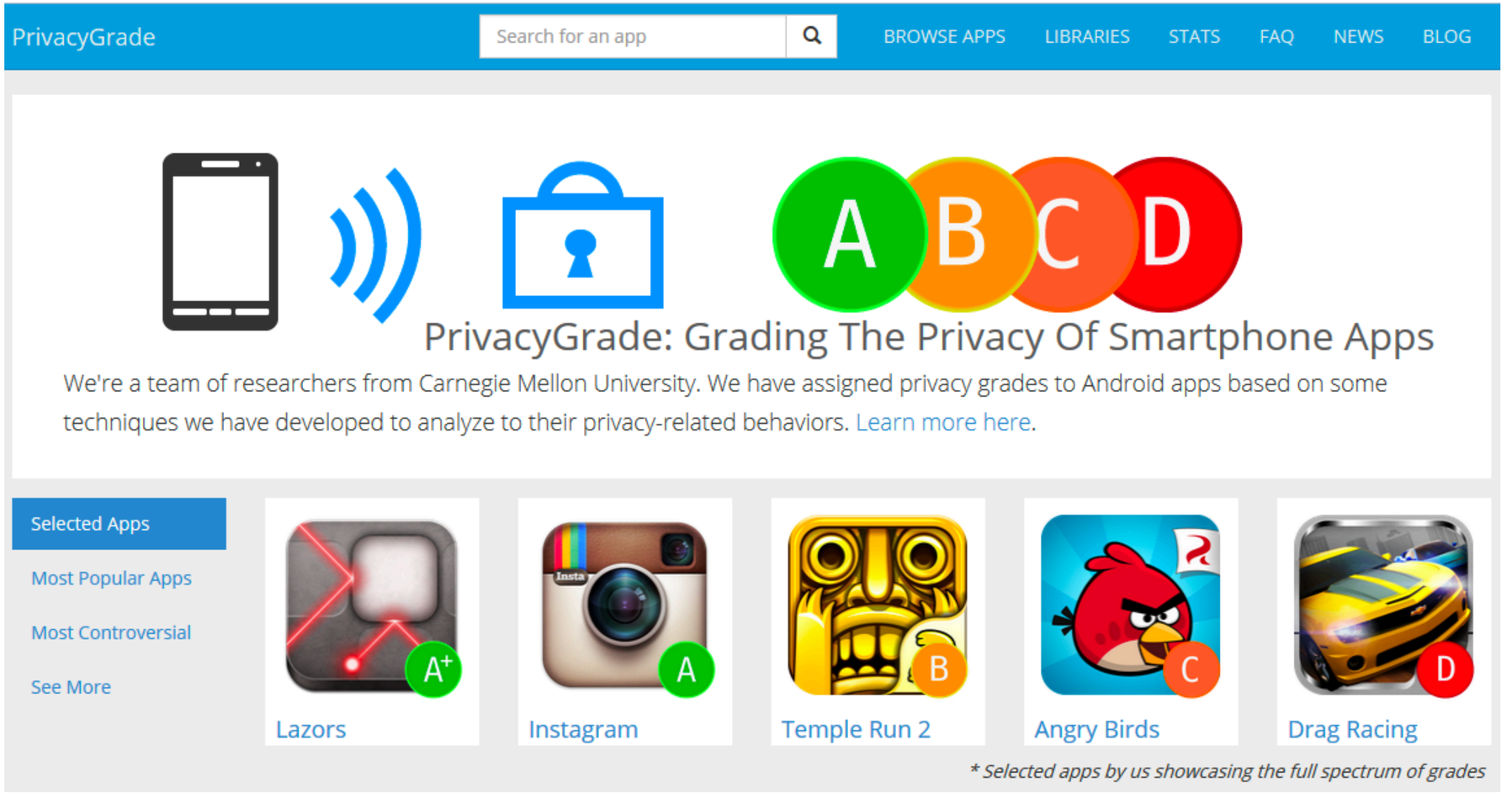
.....

Millions of small companies

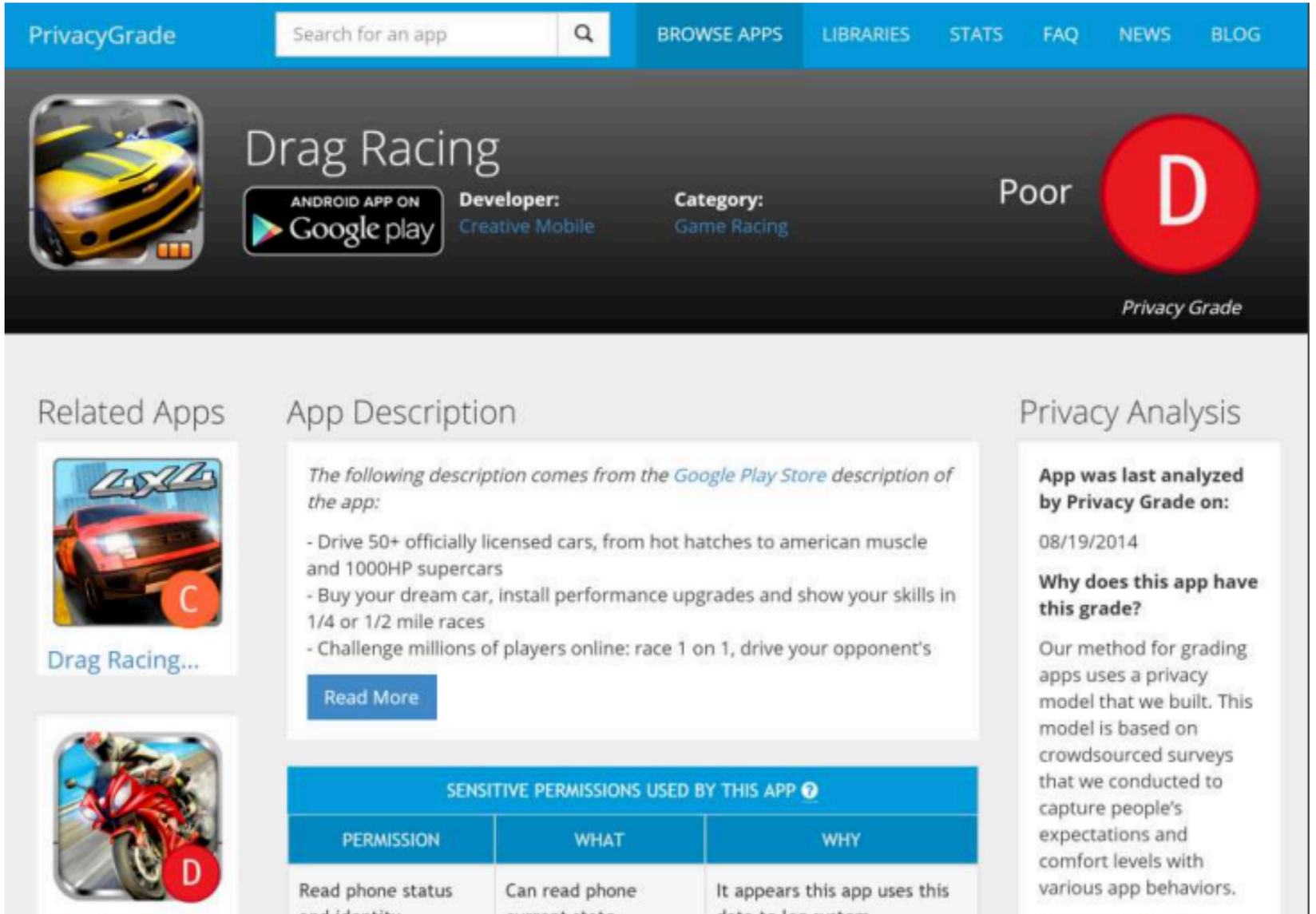
#8 Companies Get Little Pushback on Privacy

- Lack of Transparency
- Lack of Authority
- Lack of Incentives

PrivacyGrade

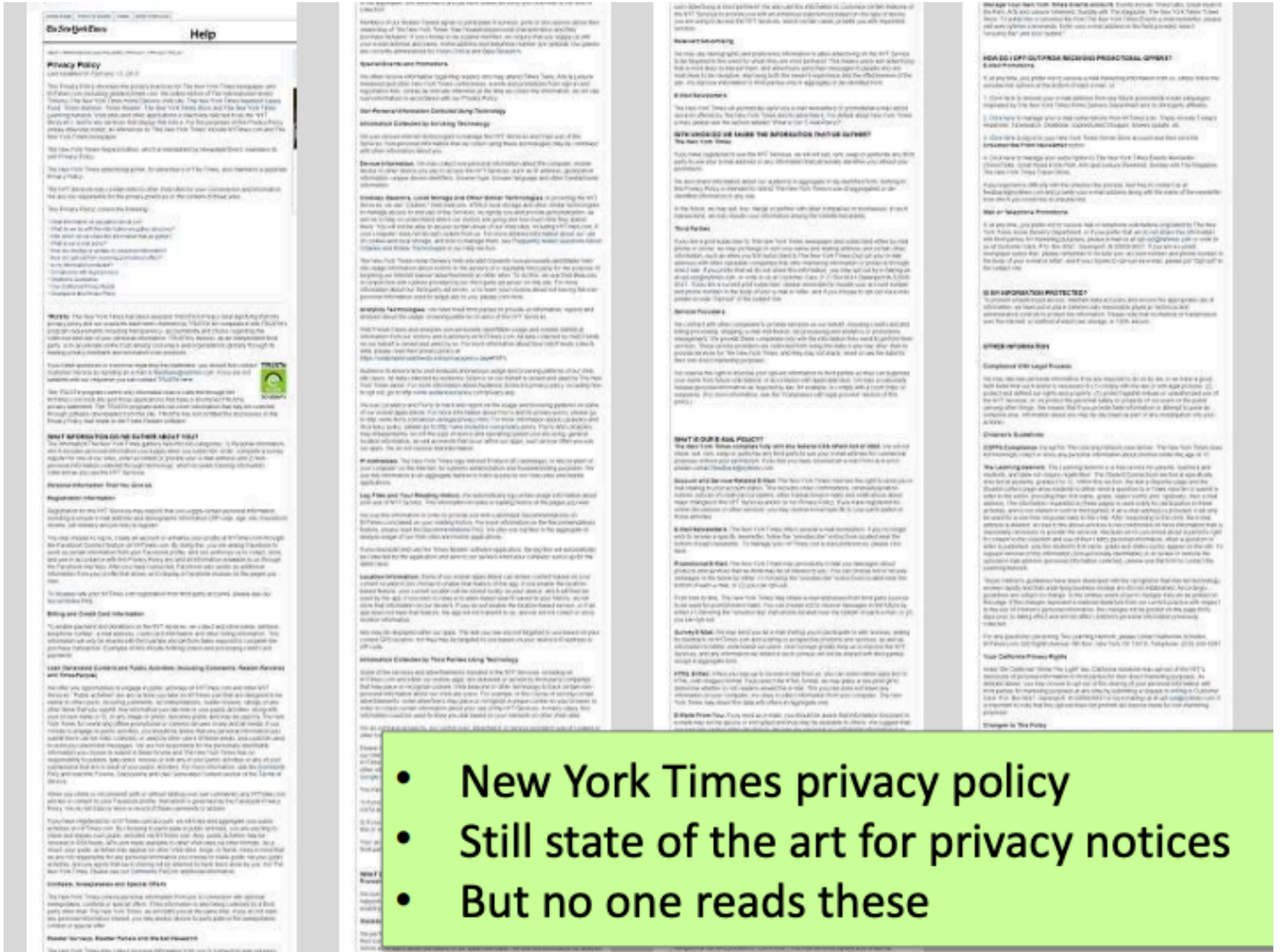


PrivacyGrade



#9 Unclear What the Right Thing To Do Is

- Even if a company wants to be privacy-sensitive,
  - It is not always clear what the right thing to do is

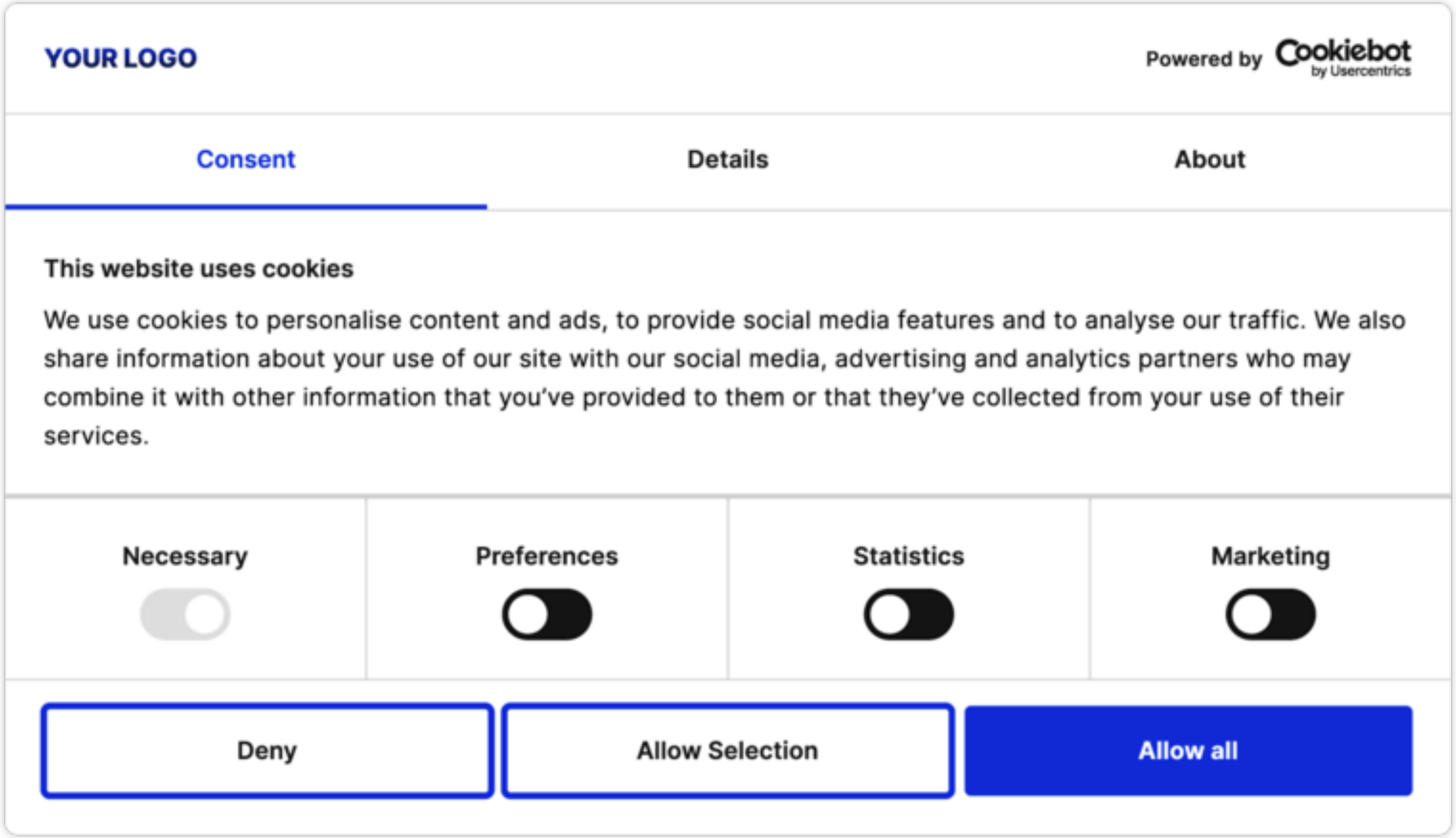


- New York Times privacy policy
- Still state of the art for privacy notices
- But no one reads these

#9 Unclear What the Right Thing To Do Is

- Even if a company wants to be privacy-sensitive,
  - It is not always clear what the right thing to do is
- For developers,
  - Best way of informing people?
  - Best way of storing data?
  - Best way of assess what is / isn't acceptable?
- For auditors,
  - How to enforce? What to push?

Consent cookie makes the problem worse.



#10 Probabilistic and Emergent Behaviors (everything else)

(a) Three samples in criminal ID photo set S_c.

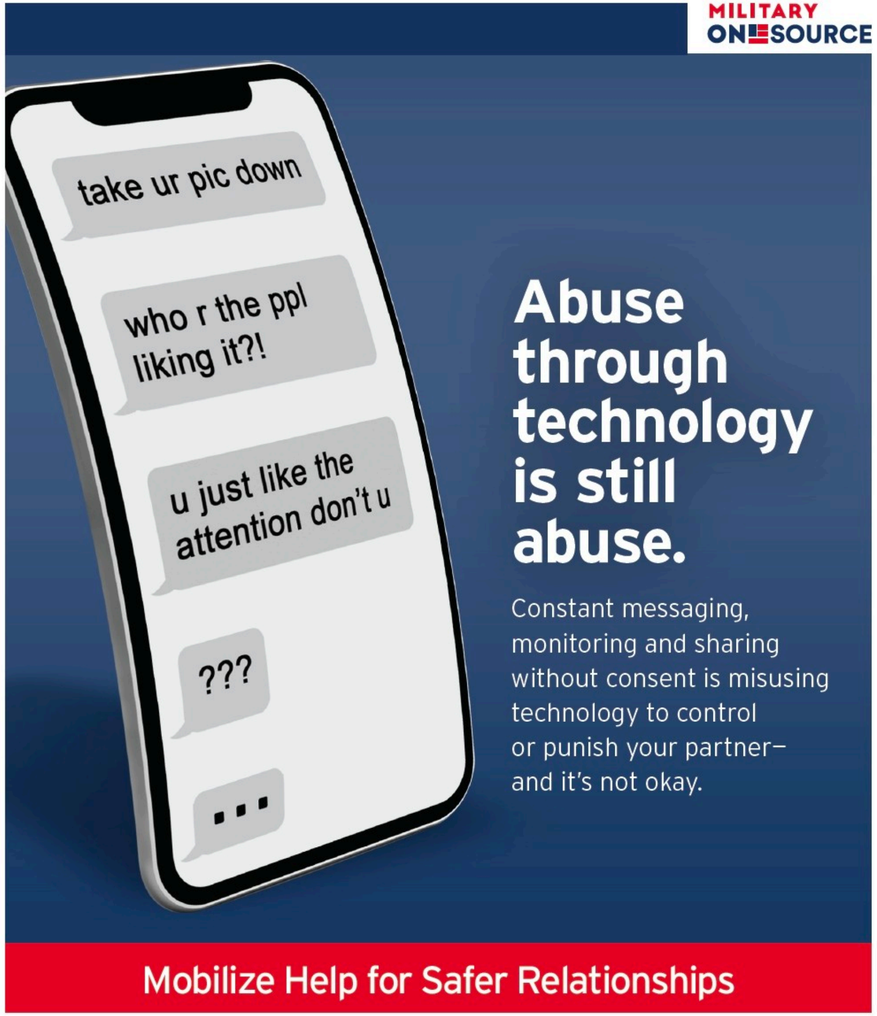
(b) Three samples in non-criminal ID photo set S_n

Figure 1. Sample ID photos in our data set.

Automated Inference on Criminality using Face Images:

95% accuracy

#10 Probabilistic and Emergent Behaviors (everything else)



Learn more about setting healthy boundaries:
www.MilitaryOneSource.mil/MobilizeHelp
Family Advocacy Program: 800-342-9647
National Domestic Violence Hotline: 800-799-7233

Summary - 10 Reasons

- #1 Privacy is a broad and fuzzy term
- #2 Technological Capabilities Rapidly Growing
- #3 Strong Incentives for Companies to Collect Data
- #4 Same Device/Data, Different Perspectives
- #5 Wide Range of Privacy Risks
- #6 Burden on End-Users Too High
- #7 Low Knowledge, Awareness, Motivation by Devs
- #8 Companies Get Little Pushback on Privacy
- #9 Unclear What the Right Thing To Do Is
- #10 Probabilistic and Emergent Behaviors