



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

Haojian Jin

Bio

Haojian Jin (<http://haojianj.in/>)

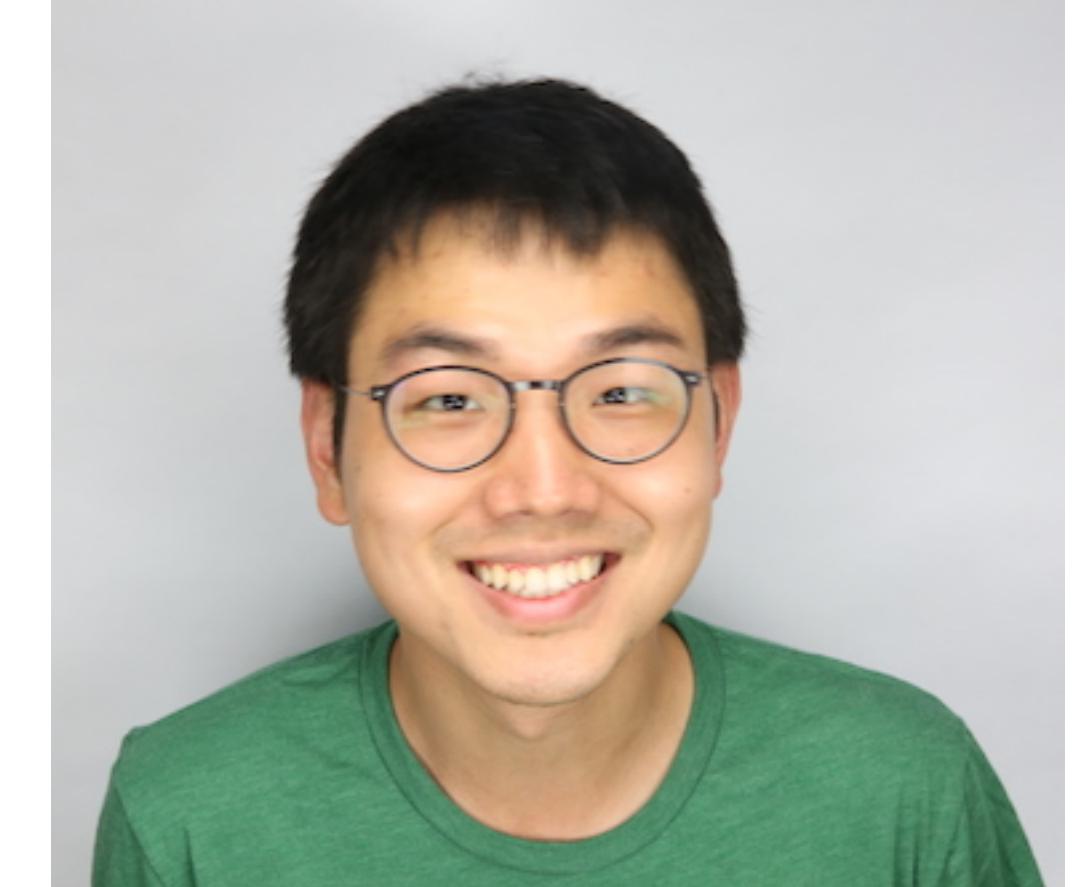
Asst. Prof @ UCSD-HDSI

HCI, Privacy, Mobile Computing

Database, Programming Language, Software Engineering

Ph.D. from CMU Human-Computer Interaction Institute

Before Ph.D.: worked at Yahoo Research, ran a startup



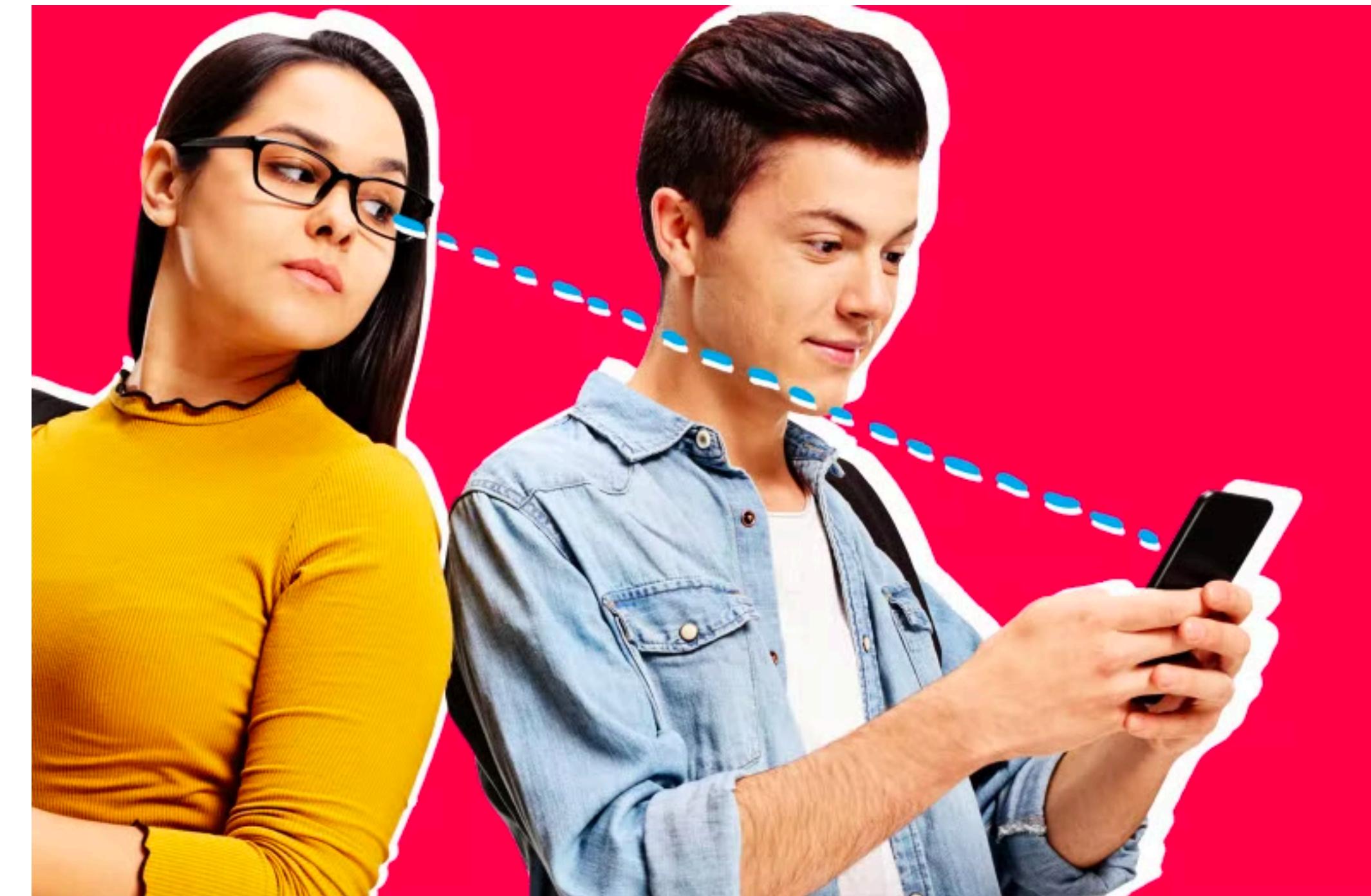
Privacy v.s. Security

- A user's bank account is hacked, and the hacker takes the user's money.
- A user's friend knows the password and logs into the user's bank account to check the spending history.
- The bank sells users' data to a third-party agency; the agency uses the user's shopping history to recommend something to that user, which may leverage the users' vulnerability.

What this course is **NOT** about?

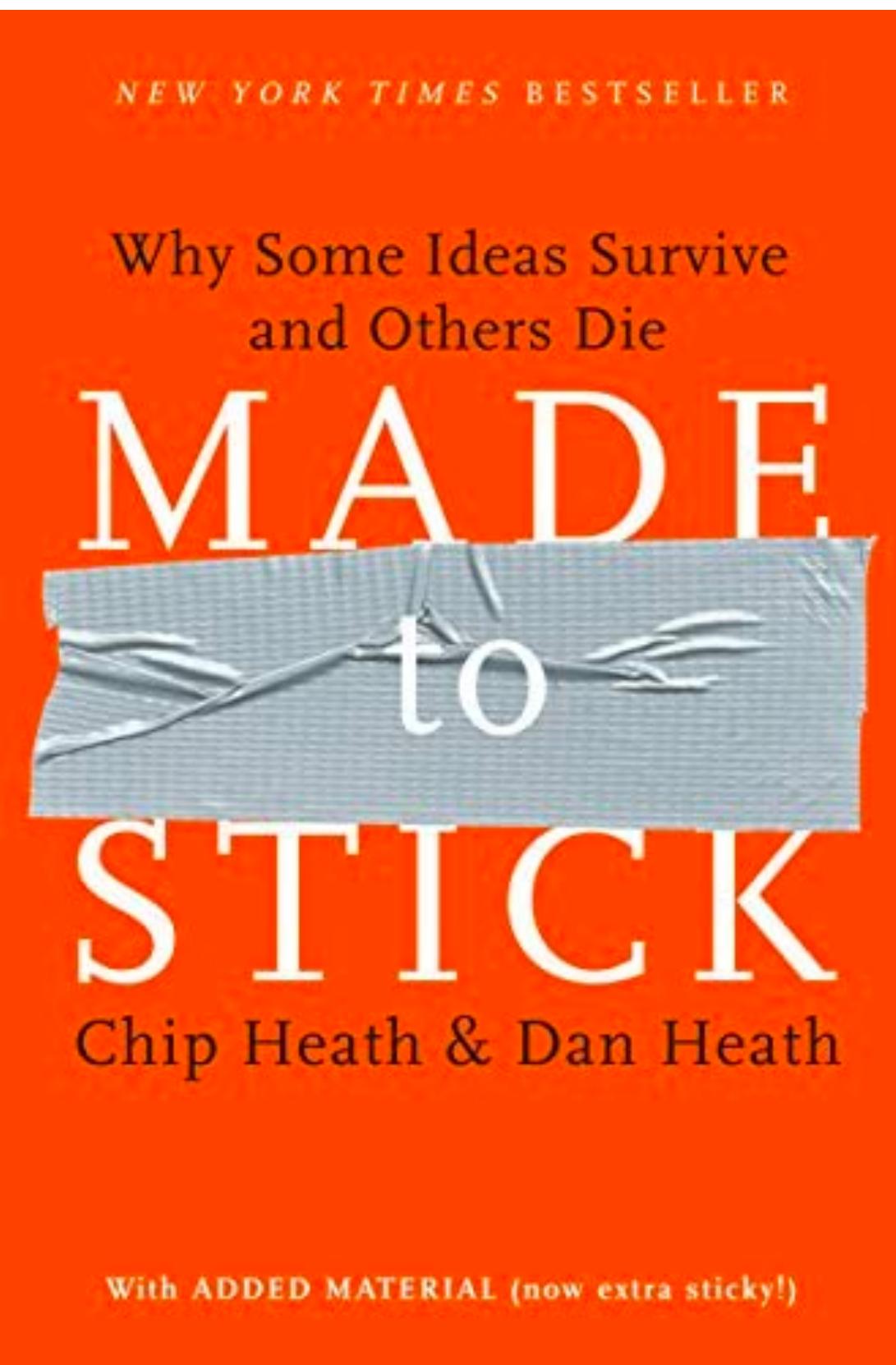


Teen's privacy



Shoulder peeking

Real-world stories!



- Understand
- Remember
- Retell

Story #1 Walmart Beer and Diaper (1988)



- Unexpected correlation:
 - Sales of diapers and beer

Forbes 1988

Story #2 Target Pregnancy (2012)



3,652,539 views | Feb 16, 2012, 11:02am

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



Kashmir Hill Former Staff
Tech

Welcome to The Not-So Private Parts where technology & privacy collide

This article is more than 2 years old.

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target [TGT +0%](#), for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.



Target has got you in its aim

Target predicted that a teenage girl might be pregnant and sent a diaper coupon to the girl.

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did.
Kashmir Hill, Forbes, 2012

A thought experiment: Data

"In our data-driven organization we collect real-time data streams and store them in our data warehouse. Our data scientists use advanced analytics and data processing in order to derive new insights."

A thought experiment: Data → Surveillance

*"In our **surveillance**-driven organization we collect real-time **surveillance** streams and store them in our **surveillance** warehouse. Our **surveillance** scientists use advanced analytics and **surveillance** processing in order to derive new insights."*

What is this course about?

- Understand range of current problems and tensions around data privacy.
- Learn why privacy is hard.
- Learn the designs/proposals/methods to address privacy problems?
 - Not just for content but context
 - Why does this design/proposal/paper exist?
 - Why do they fail?
 - Can we do sth better? (My research & your course projects!)

Why take it? A BIG GAP...



GDPR



CCPA

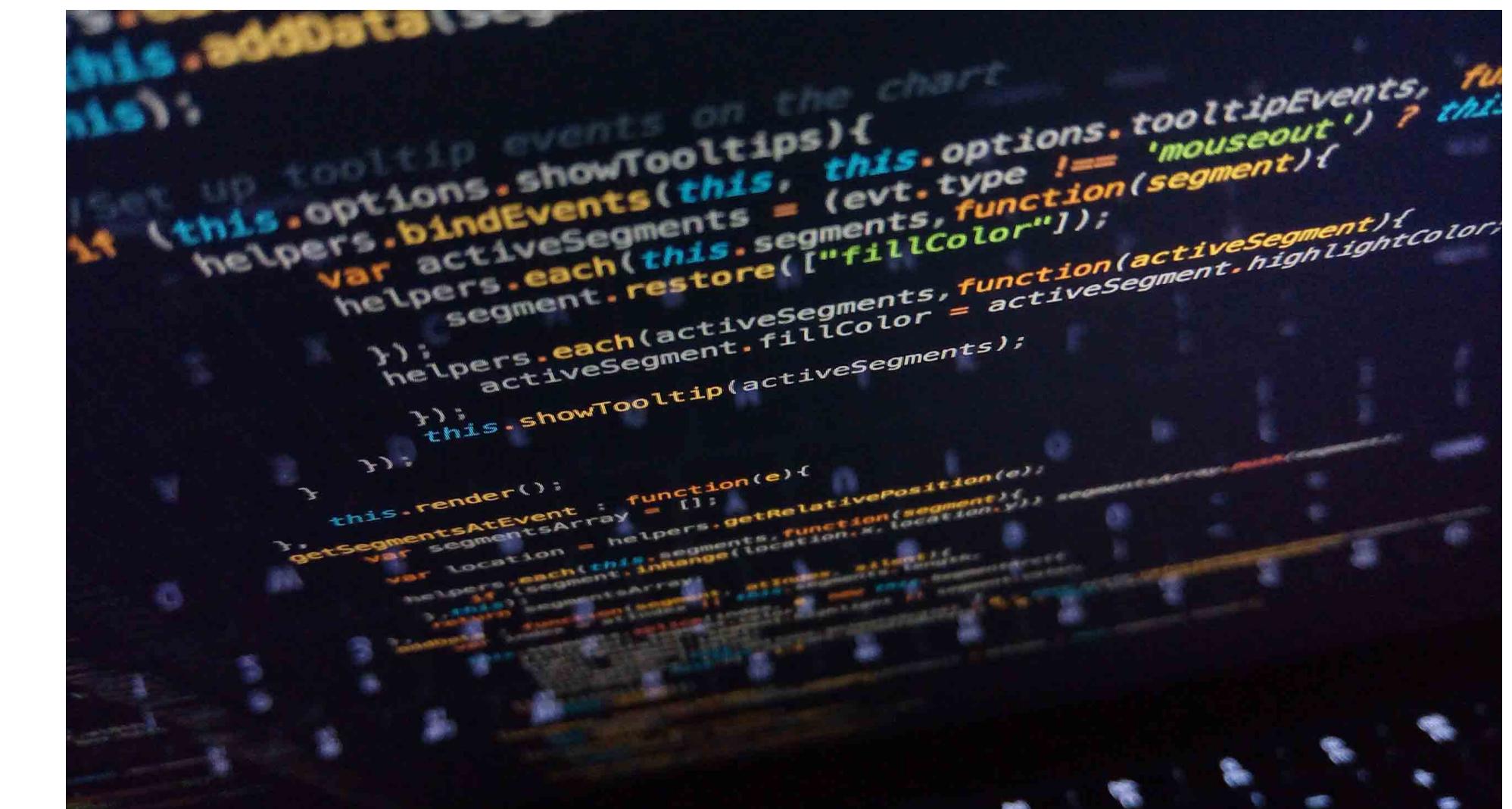
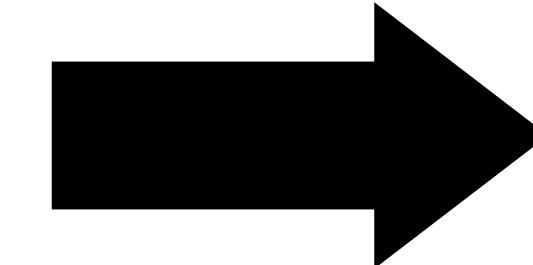


FTC



COPPA

Privacy regulations



Practice

Why take it?

Privacy is important expensive!



Public Complaints

144,376

complaints from
EU citizens

**Marriott
INTERNATIONAL**

\$123M

UK Data Protection
Authority, June 2019

Google

£50M

French Data Protection
Authority, Jan 2019

**BRITISH
AIRWAYS**

£183M

UK Data Protection
Agency, June 2019

<https://www.advisory21.com.mt/statistics-behind-the-1st-year-of-gdpr/>

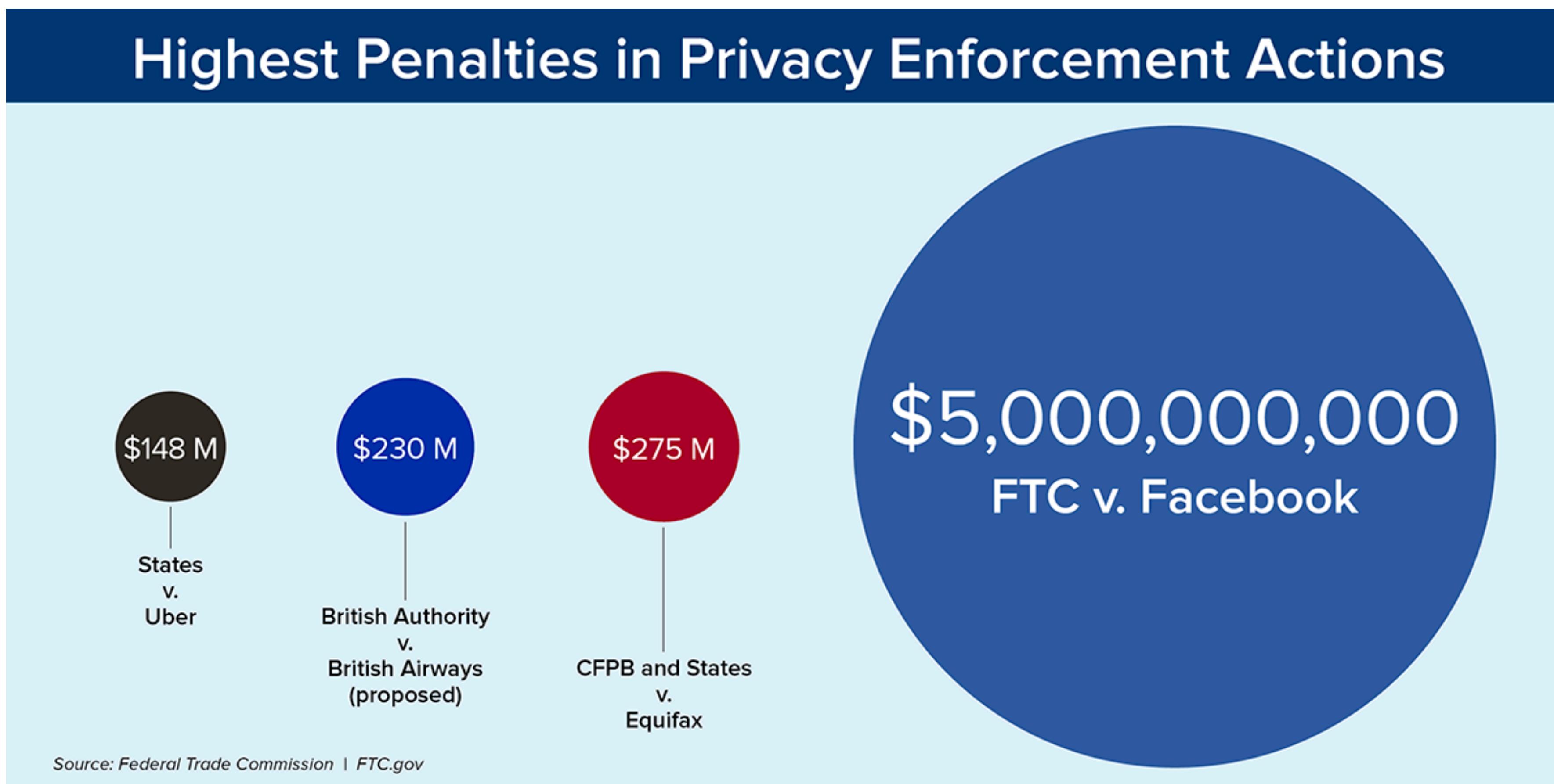
<https://www.bankinfosecurity.com/privacy-fines-total-gdpr-sanctions-reach-331-million-a-15790>

<https://idbbn.com/story/gdpr-one-year-old-and-144376-complaints/>

<https://www.bbc.com/news/business-48905907>

Why take it?

Privacy is getting more expensive!



Why take it?

It's not only about Facebook, Google, ...

Social



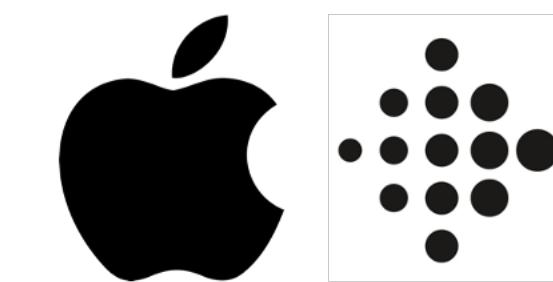
Shopping



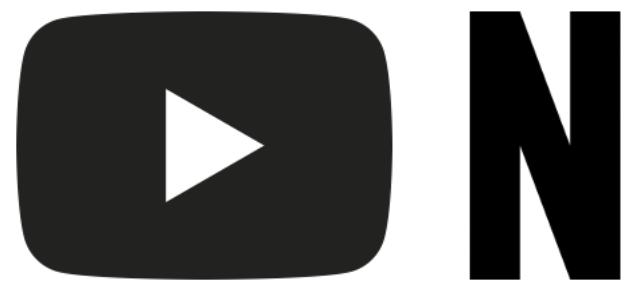
Communication



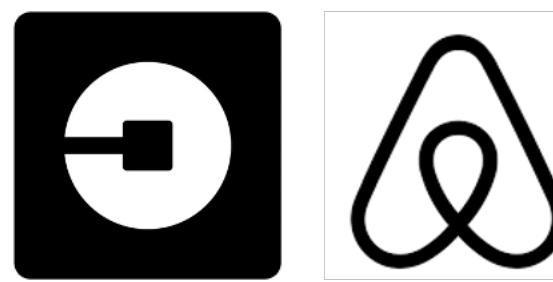
Health & Fitness



Entertainment



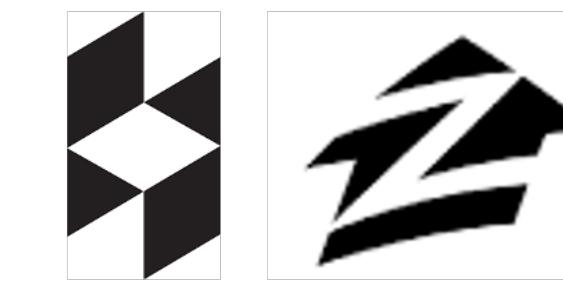
Travel



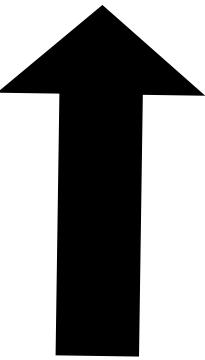
Search



Home & House



.....



**Millions of small
companies**

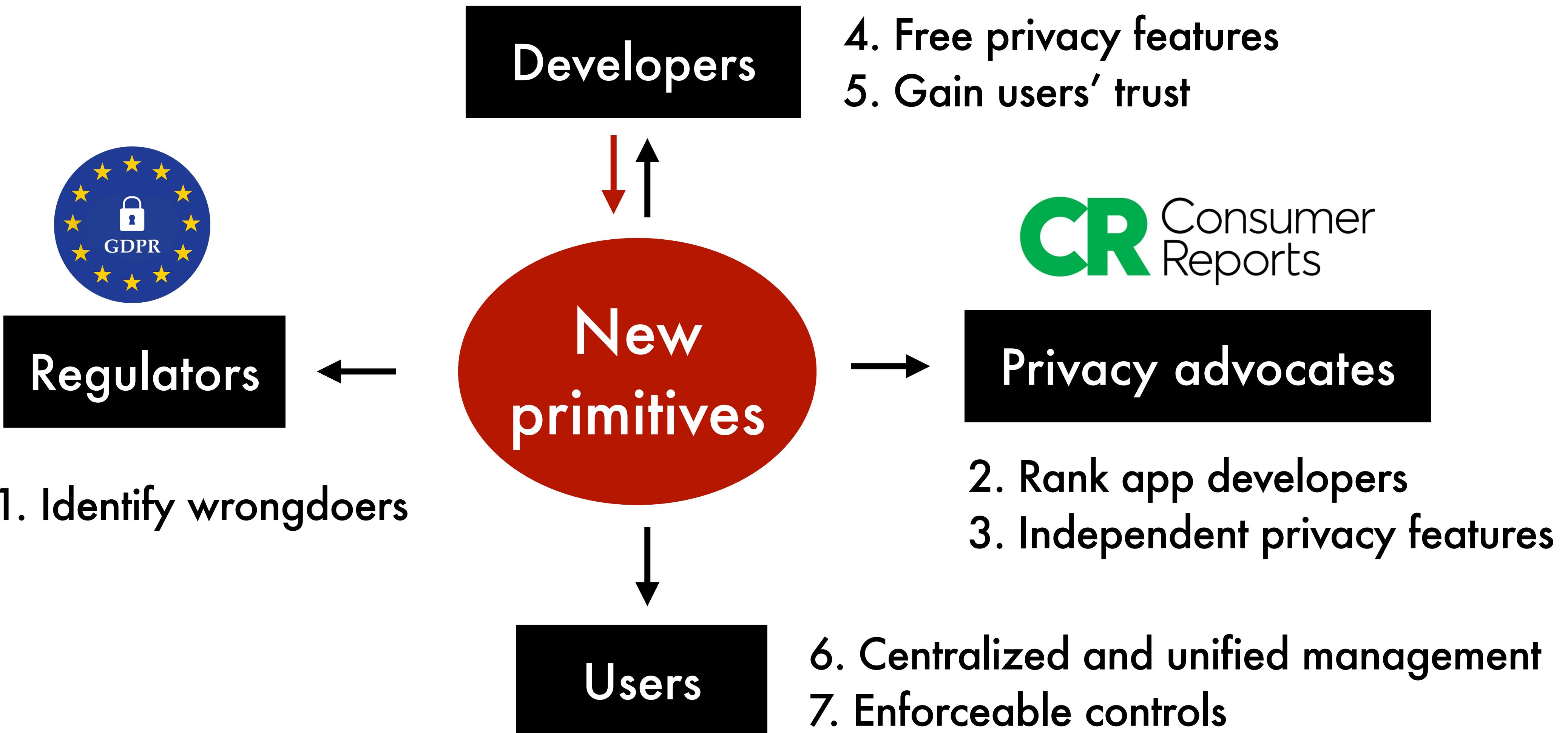
Why take it?

Building privacy-related skills takes time!

"It often take years to develop privacy-related skills and to gain an appreciation for the interplay between technical, business, human, and legal issues."

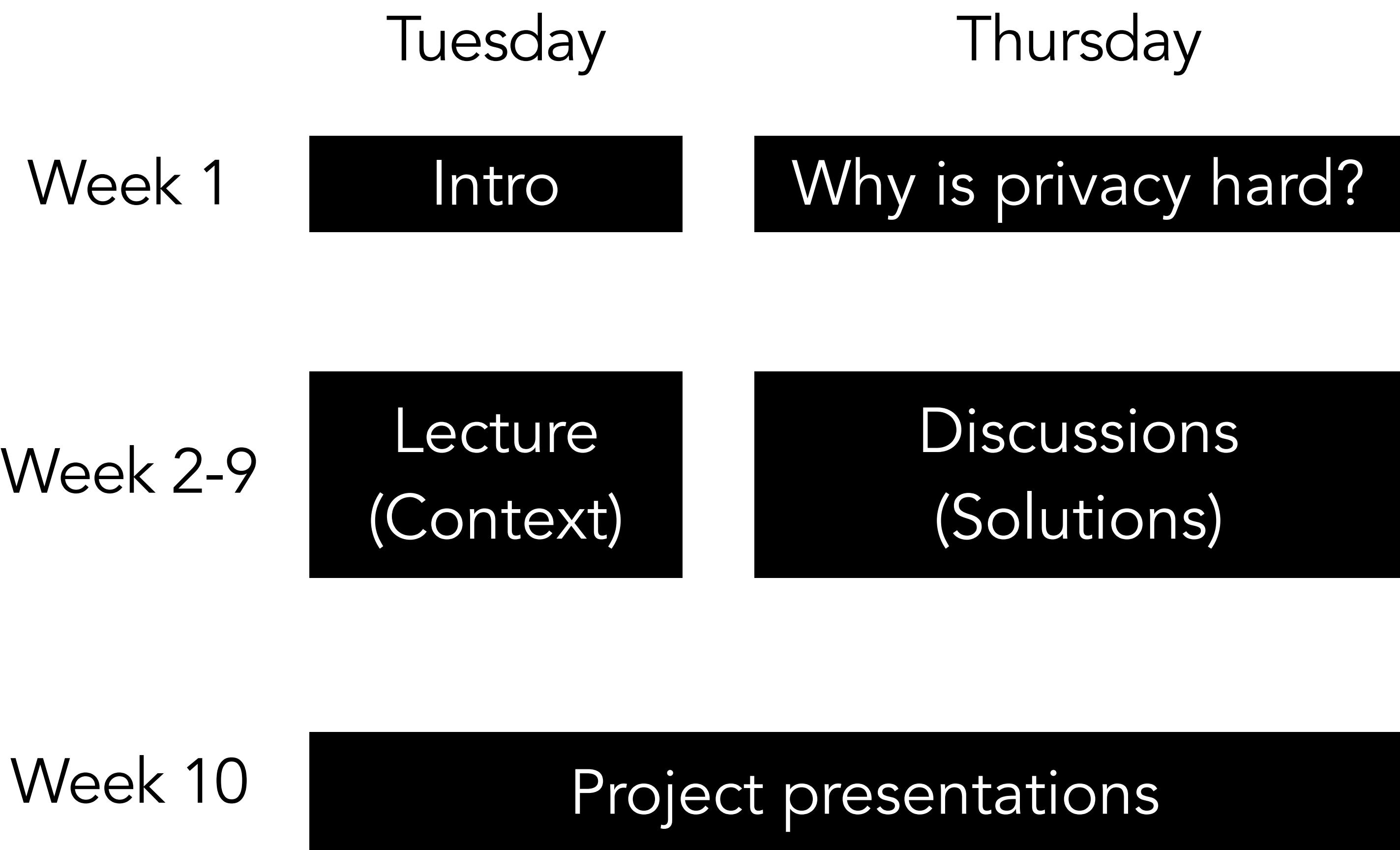
Why take it?

Get feet wet in data privacy research!



Questions?

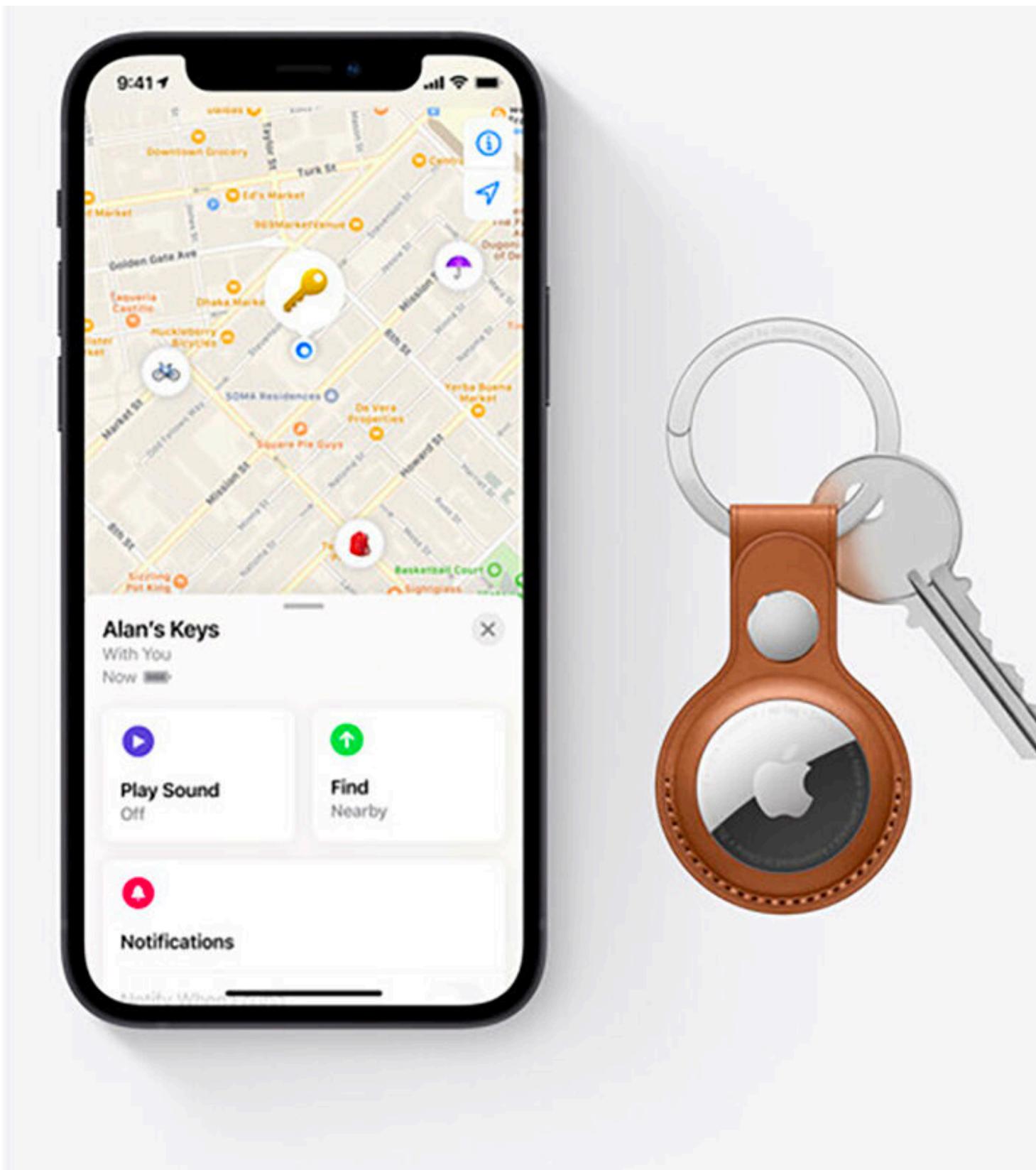
The structure of the course



The structure of the course

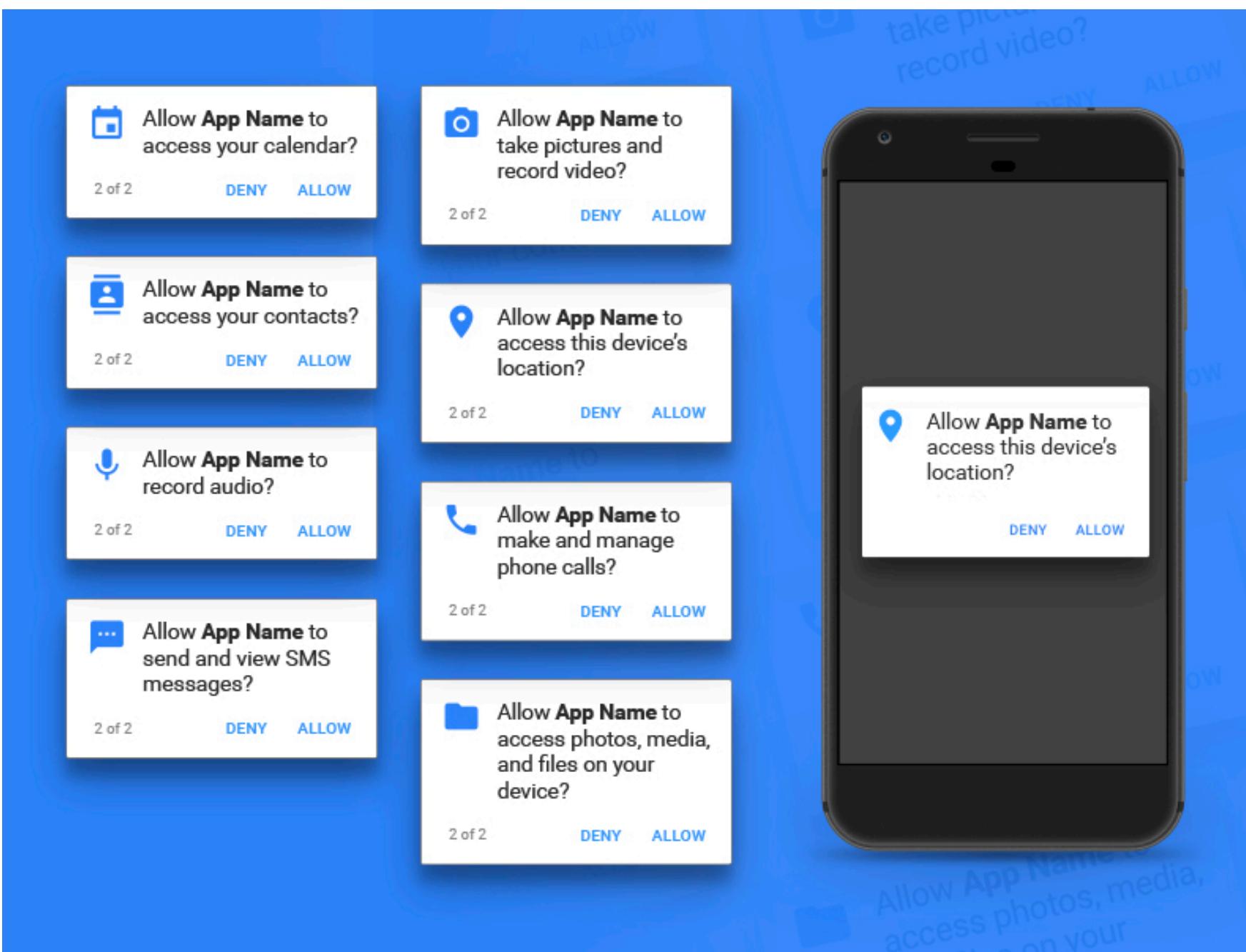
- Week 2: Apple AirTag, Contact Tracing
 - Location privacy, data collection purpose
- Week 3: Android permission, Browser cookie consent
 - Permissions, notice, control, consent, usability
- Week 4: P3P
 - Privacy policies, overcorrection, over sensing
- Week 5: Target pregnancy prediction
 - Privacy engineering, privacy expectation, contextual integrity

Week 2: Apple AirTag & contact tracing

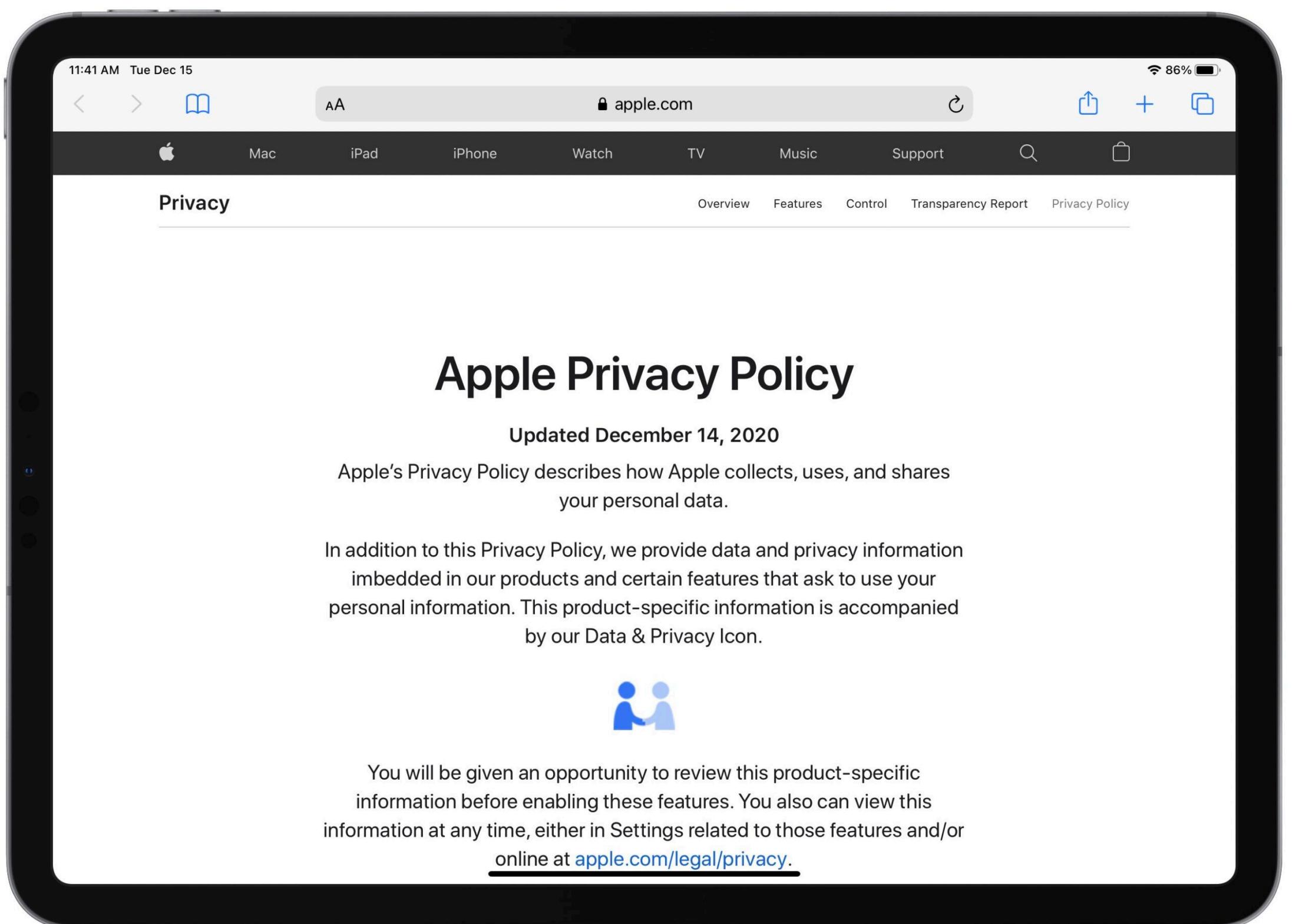


- What if you put your tag in your girlfriend's backpack?
- What if someone pick your tag?
- Does apple know everything about me?

Week 3: Android permissions & Browser cookie consent



Week 4: P3P



Data Used to Track You

The following data may be used to track you across apps and websites owned by other companies:

- Location
- Contact Info
- Browsing History
- Identifiers
- Usage Data



Data Linked to You

The following data may be collected and linked to your identity:

- Purchases
- Financial Info
- Location
- Contact Info
- Contacts
- User Content
- Search History
- Browsing History
- Identifiers
- Usage Data
- Diagnostics



Data Not Linked to You

The following data may be collected but it is not linked to your identity:

- User Content

Week 5: Target pregnancy prediction



The image shows a composite of three screenshots from the Uber app illustrating surge pricing. The first screenshot shows the main Uber interface with a map of San Francisco and a button to 'SET PICKUP LOCATION'. The second screenshot shows a 'SURGE PRICING' overlay stating 'Demand is off the charts! Fares have increased to get more Ubers on the road.' It features a large circular icon with '2.0X THE NORMAL FARE' and fare details: '\$9 MINIMUM FARE', '\$0.52 / MIN' and '\$2.60 / MILE'. The third screenshot shows another 'SURGE PRICING' overlay with a numeric keypad for confirming the fare multiple. The keypad has a grid layout:

1	2	3
4	5	6
7	8	9
PQRS	TUV	WXYZ
0		X

Text on the keypad overlay includes 'CONFIRMATION', 'Type (2.0)', 'to confirm your fare multiple.', 'MY FARE WILL BE', 'TIMES THE NORMAL FARE', and 'THIS RATE EXPIRES IN 2 MIN'.

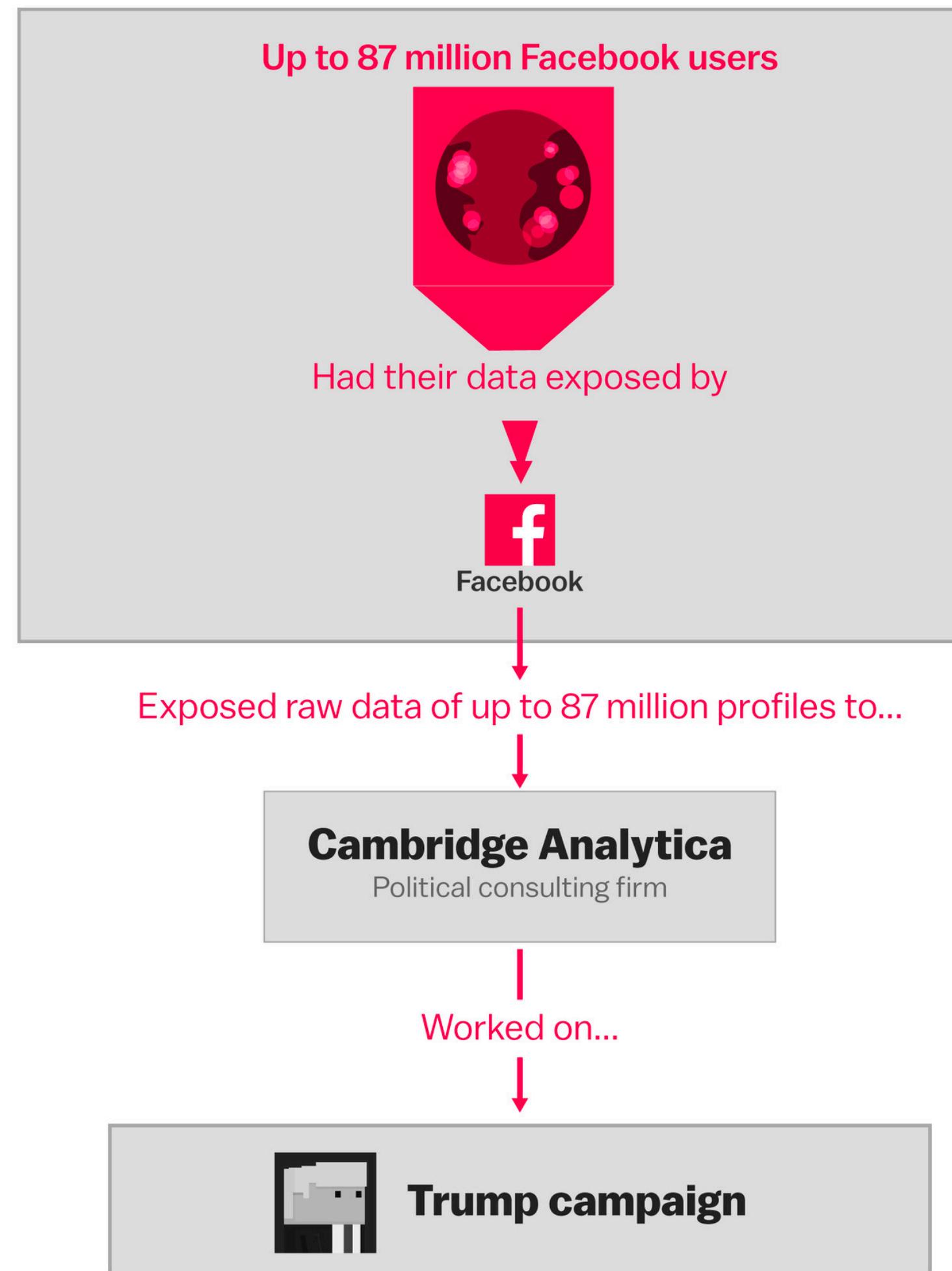
The structure of the course

- Week 6: US Census Bureau
 - Differential privacy, K-anonymity
- Week 7: Facebook cambridge analytic
 - Privacy & high stake tasks & vulnerable population
- Week 8: Big brother privacy
 - Onion routing, E2EE
- Week 9: Dark patterns
 - Agency, usability

Week 6: US Census Bureau

87 percent of all Americans could be uniquely identified using only three bits of information: ZIP code, birthdate, and sex.

Week 7: Facebook Cambridge Analytic



Week 9: Dark pattern

The screenshot shows a course landing page for "Introduction to Probability and Data". On the left, there's a sidebar with links: Syllabus, FAQs, Creators, Pricing, Ratings and Reviews, and a large green button labeled "Try for free". Below the sidebar, it says "Enroll to start full access" and has a blue "Enroll" button. A red box highlights the text "Get 6 months of access for \$235.00 USD (includes 20% discount) after the free trial or audit".

Introduction to Probability and Data

7-day Free Trial

Introduction to Probability and Data is part of the larger Statistics with R Specialization. Your 7-day free trial includes:

- ✓ **Unlimited access to all courses in the Specialization**
Watch lectures, try assignments, participate in discussion forums, and more.
- ✓ **Cancel anytime.**
No penalties - simply cancel before the trial ends if it's not right for you.
- ✓ **\$49 USD per month to continue learning after trial ends.**
Go as fast as you can - the faster you go, the more you save.
- ✓ **Certificate when you complete.**
Share on your resume, LinkedIn, and CV.

Start Free Trial

Get [6 months of access](#) for \$235.00 USD (includes 20% discount) after the free trial or [audit](#)

Learning outcome of this course

Privacy-related
data practice

Previous
solutions

Tradeoffs &
critiques

Prerequisites

- A course on database (e.g., DSC 102, DSC 202, DSC 204)
- A course on programming (e.g., DSC 200)
- The above courses could have been taken at UCSD or elsewhere.
- Industrial or substantial project experience on these topics may suffice in place of these courses. Email me if you are not sure if you satisfy the prerequisites.

Components and Grading

- Reading
 - 30 papers, around 10-20 pages of reading each week
- Reviews
 - 500 words for one paper each week.
- Student-lead discussions
 - Each student only does it once.
- Projects (best in a group of 2-3)
 - If you cannot find partner(s), let me know.

Components and Grading

- Paper reviews: 21% ($7 \times 3\%$); best 7 of 8
- Surprise in-class quizzes: 12% ($4 \times 3\%$); best 4 of 5; no-fault
- Discussion presentation: 10%.
- Project: 57%.

No other homework, midterm, or final!

Paper review

- Read three papers each week (20 - 30 pages)
- Submit reviews for one (selected by me) through Google forms.
- See deadlines on the course website.
 - 9:00 AM PST on the deadline date (the date before discussion class).
 - The forms will expire then.
 - No extension for reviews.
- Spend a few minutes looking into the background of the authors and the citations and ask yourself:
 - Why are they writing this paper? Why them? Why then? Why there?

Paper review - Grading rubrics

- Students got -1 to -10 points for grammar and clarity issues
- Students got -1 to -30 points for the review quality.

If you submit an empty review, you get 60%.

Surprise quizzes

- 20 classes; 5 in-class quizzes;
- the first 5 minutes of each class. So don't be late!
- 12% ($4 \times 3\%$); best 4 of 5; no-fault;
- No makeup quizzes.

Discussion presentation

- Each student will select a class session to be the co-discussion-leader.
- The co-leaders together will be responsible for summarizing and presenting 1 paper.
- If you find other readings that should be added, please let the professor know.
- We will assign the leaders through a Google Sheets signup (class access only), and negotiation.

Discussion instructions

- See detailed instructions in the course website.
- Expect the presentation to be about 7 minutes.
- Please prepare some slides to present in class.
- The students in the class will evaluate your presentation using Google forms.

Project

- Research project (a group of 2-3)
 - Topic due 9:00 PM Jan. 19.
- Survey project (individual project)
 - Topic due 9:00 PM Jan. 19.
- Each group need to present the project in the last week.
- Each group need to submit a final report.
 - Due 9:00 PM Mar. 18.

Project grading rubric

- Final project presentation: 40%.
- Final project report: 60%.

Things to think about...

- Pick **interesting** problems.
 - Why is this problem interesting?
- Pick problems that are achievable?
 - What resources would you need to investigate the problem?
 - You only have a quarter.
- Think about how to evaluate your work?
- I have a list of random ideas. Ask me!