How data teams can build a culture of data privacy

Josh Schwartz
Cofounder and CEO, Phaselab

Agenda

- Introduction
- Why privacy?
- Common privacy challenges
- Plenty of time for questions and discussion

A little about me

Currently, CEO and cofounder of Phaselab

ex-CTO (and, previously, Chief Data Scientist) of Chartbeat

Studied ML and computer vision as PhD student at MIT

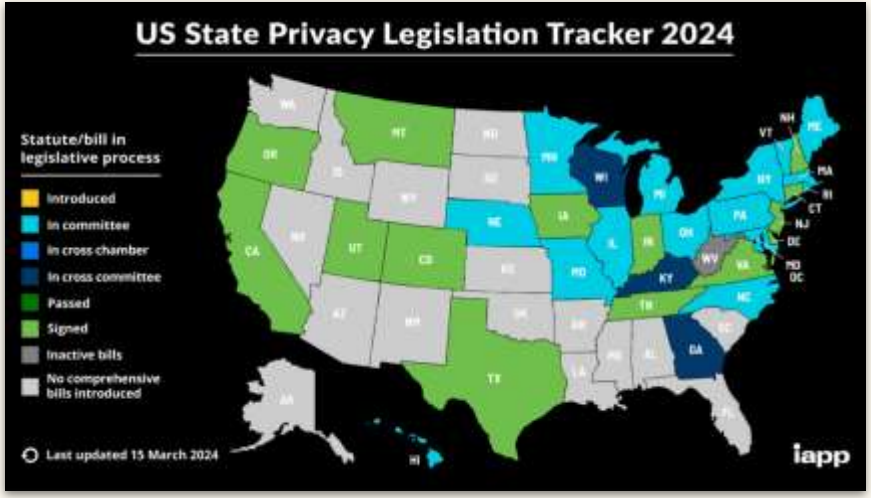
Definitely **not** a lawyer, so this isn't legal advice!

My journey in privacy

Why should *you* care about privacy?

Why privacy?

The business case for privacy



phaselab

Data teams' responsibility

Modern data infrastructure *creates* privacy risks rather than solving them

The questions that need to be answered are often too technical for your lawyers to understand, or even ask

AI makes privacy concerns 10x worse: data that was previously in cold storage is now being used as a model input

Building a culture of privacy

Creating space for privacy conversations

It’s challenging for individuals to raise privacy concerns when under pressure to ship

It’s our job as data leaders to make space for these conversations

Raise challenges as early as possible, so privacy can be designed in from the start

Privacy working groups

- Privacy is inherently cross-functional
- Needs input from legal, tech, data, marketing
- Make this group accountable for privacy — privacy can be an enormous distraction for teams
- Make legal requirements around Impact Assessments work to your advantage

Continued learning

- Privacy evolves in line with risks, laws, and consumer expectations
- Numerous certifications, conferences, and trade organizations

FUTURE OF PRIVACY FORUM

iapp

PEPR '24

Major topics in privacy that affect data teams

- Data minimization
- Maintaining consent and preferences
- Data subjects rights
- Tension between AI and privacy

Common Privacy Challenges

Data Minimization

Often the best way to avoid privacy problems is to avoid storing the data in the first place

Principle of data minimization is now enshrined in law — if you don't have a reason for storing data, you shouldn't be storing it

"I am obligated to inform you that the dog applied our document retention policy to my homework."

Minimization in practice

Retention

For each category of data, how long do we keep it?

How are our retention policies actually enforced?

Sufficiency

Do we need personal data for this use case at all?

Can the data be anonymized or obfuscated?

Data leaks

Dev environments

Local development

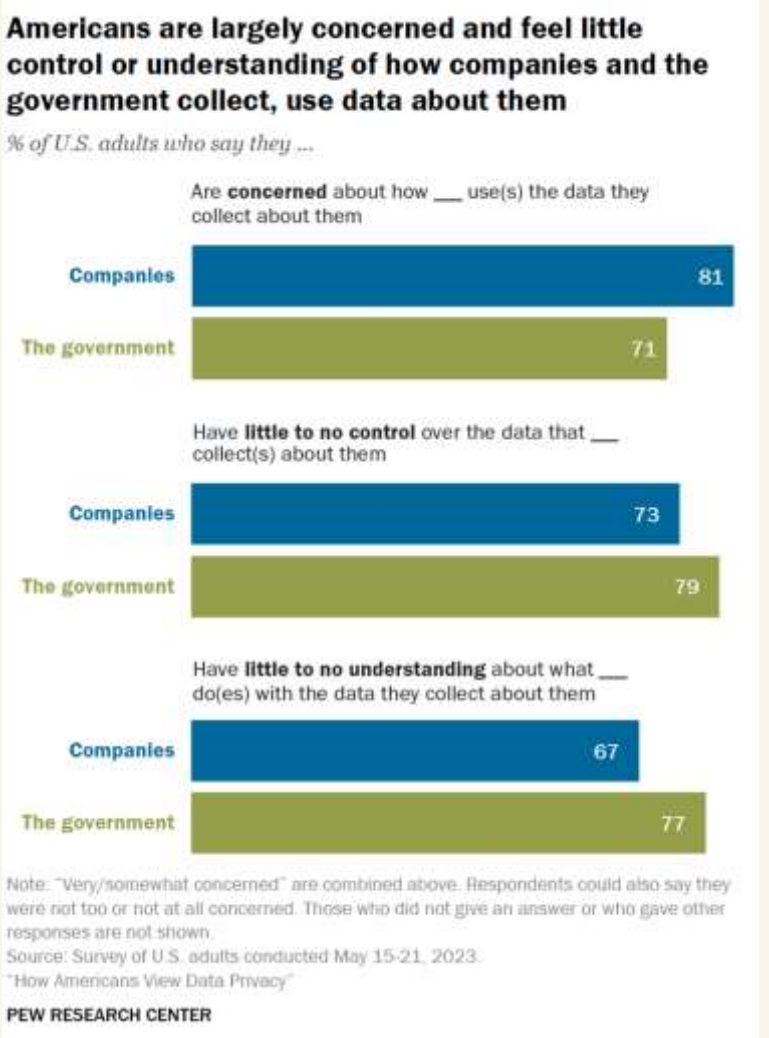
Derivative tables and models

Consent, notice and preferences

Consent is much more than just cookie banners and privacy policies

Fundamental principles:

- Consent per use case, not broad
- Consent must be opt-in, not opt-out
- Stated policies must match what you actually do



Common Privacy Challenges

Consent examples


For Release

FTC to Ban BetterHelp from Revealing Consumers' Data, Including Sensitive Mental Health Information, to Facebook and Others for Targeted Advertising

BetterHelp will be required to pay $7.8 million for deceiving consumers after promising to keep sensitive personal data private, agency says


Attorney General Bonta Announces Settlement with Sephora as Part of Ongoing Enforcement of California Consumer Privacy Act

Press Release / Attorney General Bonta Announces Settlement with Sephora as ...

Wednesday, August 24, 2022

Contact: (916) 210-6000, agpressoffice@doj.ca.gov

Marks strong second year of CCPA enforcement with update on enforcement efforts and new investigative sweep of businesses failing to process opt-out request via a user-enabled global privacy control

phaselab

Common Privacy Challenges

Data Subject Rights

The right to **be informed**

The right of **access**

The right to **correction**

The right to **deletion**

The right to **restrict processing**

The right to **data portability**

The right to **object**

Rights in relation to **automated decision making & profiling**

phaselab

Common Privacy Challenges

Data Subject Rights

The right to **be informed**

The right of **access**

The right to **correction**

The right to **deletion**

The right to **restrict processing**

The right to **data portability**

The right to **object**

Rights in relation to **automated decision making & profiling**

phaselab

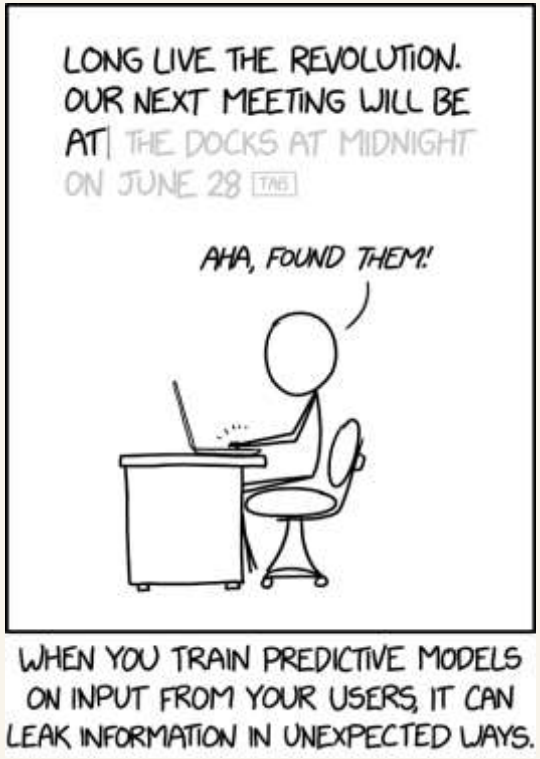
Subject rights: questions to ask

- How will I receive and apply subject rights requests (particularly deletion) across my data graph?
- What does deletion mean in the context of derived data, like rollup tables or ML models?
- How do we mitigate performance and cost implications?

AI and Privacy

Unstructured data formerly at rest is now being actively used

Tension between AI accountability and privacy



Source: https://xkcd.com/2169/

Future proofing your privacy strategy

- Privacy delivers business value in the form of trust, it’s not just a compliance exercise

Future proofing your privacy strategy

- Privacy delivers business value in the form of trust, it’s not just a compliance exercise
- Create a culture of conversation around privacy

Future proofing your privacy strategy

- Privacy delivers business value in the form of trust, it’s not just a compliance exercise
- Create a culture of conversation around privacy
- Follow principle of data minimization

Future proofing your privacy strategy

- Privacy delivers business value in the form of trust, it’s not just a compliance exercise
- Create a culture of conversation around privacy
- Follow principle of data minimization
- Plan your data architecture to make privacy policies easy to enforce

Want to learn more?

- I’m building a company around this, and I’d love to chat
- If you want to learn from the best and brightest technical folks in the space, attend PEPR (usenix.org/conference/pepr24)
- IAPP (iapp.org) is the leading Privacy industry group