Privacy in a Data-Driven World

Roxana Geambasu
Assistant Professor of Computer Science
Columbia University

https://roxanageambasu.github.io/
Example: Gmail Ads
# Example: Gmail Ads

<table>
<thead>
<tr>
<th>Email Subject &amp; Text</th>
<th>Ad Title, URL &amp; Text</th>
</tr>
</thead>
</table>
| **E1** Vacation 
I'm going on vacation to travel. | **Ad1** Ralph Lauren Online Shop 
www.ralphlauren.com 
The official Site for Ralph Lauren Apparel, Accessories & More |
| **E2** Homosexual 
Gay, lesbian, homosexual. |          |
| **E3** Pregnant 
I'm pregnant. I'm having a baby. | **Ad2** Cedars Hotel Loughborough 
www.thecedarshotel.com 
36 Bedrooms, Restaurant, Bar 
Free WiFi, Parking, Best Rates |
| **E4** Unemployed 
I'm unemployed. |          |
| **E5** Ford 
I want to buy a car, maybe a Ford. |          |
### Example: Gmail Ads

<table>
<thead>
<tr>
<th>Email Subject &amp; Text</th>
<th>Ad Title, URL &amp; Text</th>
</tr>
</thead>
</table>
| **E1** Vacation      | Ralph Lauren Online Shop  
I'm going on vacation to travel.  
www.ralphlauren.com  
The official Site for Ralph Lauren Apparel, Accessories & More |
| **E2** Homosexual    | Cedars Hotel Loughborough  
Gay, lesbian, homosexual.  
www.thecedarshotel.com  
36 Bedrooms, Restaurant, Bar  
Free WiFi, Parking, Best Rates |
| **E3** Pregnant      |                      
I'm pregnant. I'm having a baby. |
| **E4** Unemployed    |                      
I'm unemployed. |
| **E5** Ford          |                      
I want to buy a car, maybe a Ford. |
### Example: Gmail Ads

<table>
<thead>
<tr>
<th>Email Subject &amp; Text</th>
<th>Ad Title, URL &amp; Text</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vacation</strong></td>
<td><strong>Ralph Lauren Online Shop</strong></td>
</tr>
<tr>
<td>I’m going on vacation to travel.</td>
<td><a href="http://www.ralphlauren.com">www.ralphlauren.com</a> The official Site for Ralph Lauren Apparel, Accessories &amp; More</td>
</tr>
<tr>
<td><strong>Homosexual</strong></td>
<td><strong>Cedars Hotel Loughborough</strong></td>
</tr>
<tr>
<td>Gay, lesbian, homosexual.</td>
<td><a href="http://www.thecedarshotel.com">www.thecedarshotel.com</a> 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates</td>
</tr>
<tr>
<td><strong>Pregnant</strong></td>
<td></td>
</tr>
<tr>
<td>I’m pregnant. I’m having a baby.</td>
<td></td>
</tr>
<tr>
<td><strong>Unemployed</strong></td>
<td></td>
</tr>
<tr>
<td>I’m unemployed.</td>
<td></td>
</tr>
<tr>
<td><strong>Ford</strong></td>
<td></td>
</tr>
<tr>
<td>I want to buy a car, maybe a Ford.</td>
<td></td>
</tr>
</tbody>
</table>
### Example: Gmail Ads

<table>
<thead>
<tr>
<th>Email subject &amp; text</th>
<th>Ad title, url &amp; text</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vacation</strong></td>
<td><strong>Ralph Lauren Online Shop</strong></td>
</tr>
<tr>
<td>I'm going on vacation to travel.</td>
<td><a href="http://www.ralphlauren.com">www.ralphlauren.com</a> The official Site for Ralph Lauren Apparel, Accessories &amp; More</td>
</tr>
<tr>
<td><strong>Homosexual</strong></td>
<td><strong>Cedars Hotel Loughborough</strong></td>
</tr>
<tr>
<td>Gay, lesbian, homosexual.</td>
<td><a href="http://www.thecedarhotel.com">www.thecedarhotel.com</a> 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates</td>
</tr>
<tr>
<td><strong>Pregnant</strong></td>
<td></td>
</tr>
<tr>
<td>I'm pregnant. I'm having a baby.</td>
<td></td>
</tr>
<tr>
<td><strong>Unemployed</strong></td>
<td></td>
</tr>
<tr>
<td>I'm unemployed.</td>
<td></td>
</tr>
<tr>
<td><strong>Ford</strong></td>
<td></td>
</tr>
<tr>
<td>I want to buy a car, maybe a Ford.</td>
<td></td>
</tr>
</tbody>
</table>
## Example: Gmail Ads

<table>
<thead>
<tr>
<th>Email subject &amp; text</th>
<th>Ad title, url &amp; text</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vacation</strong></td>
<td><strong>Ralph Lauren Online Shop</strong></td>
</tr>
<tr>
<td>I'm going on vacation to travel.</td>
<td><a href="http://www.ralphlauren.com">www.ralphlauren.com</a> The official Site for Ralph Lauren Apparel, Accessories &amp; More</td>
</tr>
<tr>
<td><strong>Homosexual</strong></td>
<td><strong>Cedars Hotel Loughborough</strong></td>
</tr>
<tr>
<td>Gay, lesbian, homosexual.</td>
<td><a href="http://www.thecedarshotel.com">www.thecedarshotel.com</a> 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates</td>
</tr>
<tr>
<td><strong>Pregnant</strong></td>
<td></td>
</tr>
<tr>
<td>I'm pregnant. I'm having a baby.</td>
<td></td>
</tr>
<tr>
<td><strong>Unemployed</strong></td>
<td></td>
</tr>
<tr>
<td>I'm unemployed.</td>
<td></td>
</tr>
<tr>
<td><strong>Ford</strong></td>
<td></td>
</tr>
<tr>
<td>I want to buy a car, maybe a Ford.</td>
<td></td>
</tr>
</tbody>
</table>
It’s not just Gmail...

Did you know?

• Data brokers can tell when you're sick, tired and depressed -- and sell this information. [CNN ’14]

• Google Apps for Ed used institutional emails to target ads in personal accounts. [SafeGov’14]

• Credit companies are looking into using Facebook data to decide loans. [CNN’13]
The data-driven web

- The web is a **complex and opaque ecosystem** driven by massive collection and monetization of personal data.

- Who has what data?
- What’s it used for?
- Are the uses good or bad for us?
- End-users, privacy watchdogs (e.g., FTC) are equally blind.
My research

1. Build transparency tools that increase users’ awareness and society’s oversight over how apps use personal data:
   - **Sunlight**: reveals the causes of targeting [CCS’15].
   - **XRay**: reveals targeting through correlation [USENIX Sec’14].
   - **Pebbles**: reveals how mobile apps manage persistent data [OSDI’14].
My research

1. Build transparency tools that increase users’ awareness and society’s oversight over how apps use personal data:
   - **Sunlight**: reveals the causes of targeting [CCS’15].
   - **XRay**: reveals targeting through correlation [USENIX Sec’14].
   - **Pebbles**: reveals how mobile apps manage persistent data [OSDI’14].

2. Build development abstractions and tools that facilitate construction of privacy-preserving apps:
   - **FairTest**: unit tests for fairness [under review].
   - **CleanOS**: privacy-mindful mobile operating system [OSDI’12].
   - **Pyramid**: minimizing data exposure in data-driven apps [ramping up].
My research

1. Build **transparency tools** that increase users’ awareness and society’s oversight over how apps use personal data:
   - **Sunlight**: reveals the causes of targeting [CCS’15].
   - **XRay**: reveals targeting through correlation [USENIX Sec’14].
   - **Pebbles**: reveals how mobile apps manage persistent data [OSDI’14].

2. Build **development abstractions and tools** that facilitate construction of privacy-preserving apps:
   - **FairTest**: unit tests for fairness [under review].
   - **CleanOS**: privacy-mindful mobile operating system [OSDI’12].
   - **Pyramid**: minimizing data exposure in data-driven apps [ramping up].
my students:

- Vaggelis Atlidakis
- Mathias Lecuyer
- Riley Spahn
- Yannis Spiliopoulos

some of my collaborators:

- Augustin Chaintreau (Columbia)
- Daniel Hsu (Columbia)
- Jean-Pierre Hubaux (EPFL)
- Ari Juels (Cornell Tech)
Sunlight: transparency for the data-driven web.

[CCS’15]
Sunlight

- Generic and broadly applicable system that detects personal data use for targeting and personalization.
  - Reveals which data (e.g., emails) triggers which outputs (e.g., ads).

- Key idea: correlate inputs with outputs based on observations from profiles with differentiated inputs.

- Sunlight is precise, scalable, and works with many services.
  - We tested it for Gmail ads, ads on arbitrary websites, recommendations on Amazon & YouTube, prices in travel websites.
**Example**

**email subject & text**

<table>
<thead>
<tr>
<th>Email</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>Vacation</td>
</tr>
<tr>
<td></td>
<td>I'm going on vacation to travel.</td>
</tr>
<tr>
<td>E2</td>
<td>Homosexual</td>
</tr>
<tr>
<td></td>
<td>Gay, lesbian, homosexual.</td>
</tr>
<tr>
<td>E3</td>
<td>Pregnant</td>
</tr>
<tr>
<td></td>
<td>I'm pregnant. I'm having a baby.</td>
</tr>
<tr>
<td>E4</td>
<td>Unemployed</td>
</tr>
<tr>
<td></td>
<td>I'm unemployed.</td>
</tr>
<tr>
<td>E5</td>
<td>Ford</td>
</tr>
<tr>
<td></td>
<td>I want to buy a car, maybe a Ford.</td>
</tr>
</tbody>
</table>

**ad title, url & text**

<table>
<thead>
<tr>
<th>Ad</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ad1</td>
<td>Ralph Lauren Online Shop</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.ralphlauren.com">www.ralphlauren.com</a></td>
</tr>
<tr>
<td></td>
<td>The official Site for Ralph Lauren Apparel, Acccessories &amp; More</td>
</tr>
<tr>
<td>Ad2</td>
<td>Cedars Hotel Loughborough</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.thecedarshotel.com">www.thecedarshotel.com</a></td>
</tr>
<tr>
<td></td>
<td>36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates</td>
</tr>
</tbody>
</table>
### Example

<table>
<thead>
<tr>
<th>Email subject &amp; text</th>
<th>Ad title, url &amp; text</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vacation</strong></td>
<td><strong>Ralph Lauren Online Shop</strong></td>
</tr>
<tr>
<td>I’m going on vacation to travel.</td>
<td><a href="http://www.ralphlauren.com">www.ralphlauren.com</a></td>
</tr>
<tr>
<td><strong>Homosexual</strong></td>
<td>The official Site for Ralph Lauren Apparel, Accessories &amp; More</td>
</tr>
<tr>
<td>Gay, lesbian, homosexual.</td>
<td></td>
</tr>
<tr>
<td><strong>Pregnant</strong></td>
<td>Ad1</td>
</tr>
<tr>
<td>I’m pregnant. I’m having a baby.</td>
<td></td>
</tr>
</tbody>
</table>
Example

main account

<table>
<thead>
<tr>
<th>E1</th>
<th>Ad1</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2</td>
<td></td>
</tr>
<tr>
<td>E3</td>
<td></td>
</tr>
</tbody>
</table>
Example

main account

E1
E2
E3

Ad1

shadow account 1

E1
E2

shadow account 2

E1
E3

shadow account 3

E2
E3
Example

main account

E1
E2
E3

Ad1

shadow account 1

E1
E2

shadow account 2

E1
E3

Ad1

shadow account 3

E2
E3

Ad1
Example

main account

E1
E2
E3

Ad1

shadow account 1

E1
E2

shadow account 2

E1
E3

Ad1

shadow account 3

E2
E3

Ad1

targeting prediction:

E3

Ad1
data collection: service-specific, with browser automation

main account

E1
E2
E3

Ad1

shadow account 1

E1
E2

shadow account 2

E1
E3
Ad1

shadow account 3

E2
E3

Ad1

targeting analysis: service-agnostic, with Sunlight

targeting prediction:
Transparency solutions

end-users, privacy watchdogs (e.g., FTC, journalists)

GmailAd-Observatory

AdsOnWeb-Observatory

AMZN,Youtube recommendations

transparency tools (built by us, others)

Sunlight
(generic, scalable, and justifiable targeting detection)

input/output observations

targeting predictions {inputs->output}

transparency infrastructures
Sunlight talk

- Overview
- Design
- Evaluation
- Use cases
Design goals

- **Generic and broadly applicable targeting detection**
  - We assume that a small set of inputs is used to produce each output. Our goal is to discover the *correct* input combination.

- **Precise and justifiable targeting predictions**
  - Targeting predictions must be statistically justified. Our goal is to detect as many *true* predictions as possible.

- **Scalable in number of inputs and outputs**
  - Detect targeting of many outputs on many inputs w/ limited resources.
The scalability challenge

- To detect targeting on combinations of the inputs, will we need shadow profiles for all combinations???
Scalable targeting detection

- **Theorem:** Under sparsity assumptions, for any $\varepsilon > 0$ there exists an algorithm that requires $C \times \log(N)$ accounts to correctly identify the inputs of a targeted output with probability $(1 - \varepsilon)$. $N$ is the number of inputs.

- Key insight: rely on sparsity properties (like compressed sensing).

- We incorporate several sparse detection algorithms:
  - **Set intersection** -- simple, not robust
  - **Sparse regressions (Lasso)** -- well established, robust
Justifiable targeting predictions

- Sparse algorithms only guarantee asymptotic correctness of the targeting predictions.
- We need **correctness assessment** for each targeting prediction.

- Solution: **hypothesis testing**.
  - Provides quantification of statistical significance of each targeting association (a p-value).
  - p-value gives knob for precision/recall tradeoff.
Architecture

Transparency tool (e.g., GmailAdObservatory)

input/output observations

justifiable targeting predictions & p-values

Sunlight
Architecture

Input/output observations

Transparent tool (e.g., Gmail/Ad Observatory)

justifiable targeting predictions & p-values

Split observations

training set

putative targeting predictions

Scalable Targeting Prediction

Testing set

Prediction Hypothesis Testing

Prediction Filtering

Multiple Test Correction

targeting predictions & p-values

filtered targeting predictions

Sunlight
Architecture

Sunlight

Split observations

Training set

Testing set

Scalable Targeting Prediction

Putative targeting predictions

Prediction Filtering

Prediction Hypothesis Testing

Targeting predictions & p-values

Multiple Test Correction

Targeting predictions & p-values

Input/output observations

Transparency tool (e.g., Gmail/Ad Observatory)

Justifiable targeting predictions & p-values
Architecture

Sunlight

input/output observations

Split observations

training set

testing set

putative targeting predictions

Multiple Test Correction

targeting predictions & p-values

Prediction Hypothesis Testing

filtered targeting predictions

Prediction Filtering

justifiable targeting predictions & p-values

Transparency tool (e.g., GmailAdObservatory)
Architecture

input/output observations → Split observations → training set → Scalable Targeting Prediction

justifiable targeting predictions & p-values → Multiple Test Correction → targeting predictions & p-values → Prediction Hypothesis Testing → filtered targeting predictions → Prediction Filtering
Architecture

- Transparency tool (e.g., Gmail/AdObservatory)
  - input/output observations
  - justifiable targeting predictions & p-values

- Split observations
- training set
- testing set
- Scalable Targeting Prediction
- putative targeting predictions
- Multiple Test Correction
  - targeting predictions & p-values
- Prediction Hypothesis Testing
- filtered targeting predictions
- Prediction Filtering
What we get in the end

- If during data collection we randomly assign our inputs independently of any other variable, Sunlight’s associations will have a causal interpretation (not just correlation).

- However, Sunlight cannot explain how this targeting happens.
  - E.g.: What player in the ecosystem is responsible? Is it a human intervention or an algorithmic decision?
Sunlight talk

- Overview
- Design
- Evaluation
- Use cases
## Datasets

<table>
<thead>
<tr>
<th>Workload</th>
<th>Profiles</th>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gmail (one day)</td>
<td>119</td>
<td>327</td>
<td>4099</td>
</tr>
<tr>
<td>Website</td>
<td>200</td>
<td>84</td>
<td>4867</td>
</tr>
<tr>
<td>Website-large</td>
<td>798</td>
<td>263</td>
<td>19808</td>
</tr>
<tr>
<td>YouTube</td>
<td>45</td>
<td>64</td>
<td>308</td>
</tr>
<tr>
<td>Amazon</td>
<td>51</td>
<td>61</td>
<td>2593</td>
</tr>
</tbody>
</table>
Targeting prediction precision

We developed two methodologies:

1. **Manual assessments** of how “believable” are our low-p-value predictions (<0.05).
   - We observed 100% precision for smaller experiments and 95%-96% precision for larger experiments. Despite potential for confirmation bias, this is in line with expectation at p-value < 0.05.

2. Assess the **quality of targeting predictions**.
   - If we conclude that E3->Ad1, we should be able to use E3’s presence in a shadow account to accurately guess whether Ad1 appears in that account.
Quality of targeting predictions

Y: Proportion of ad appearances that were correctly guessed to be present in a shadow account.
Quality of targeting predictions

Y: Proportion of success when guessing if an ad will be present in a shadow account.
Targeting prediction recall

- We found recall impossible to quantify manually.
  - Too many outputs, too many input possibilities, too error prone.

- We developed this methodology:
  - Inspected ads for which Sunlight had some evidence they were being targeted, but for which correction spoiled their p-values.
  - This methodology revealed a precision-recall tradeoff at scale due to correction.
Precision/recall tradeoff

p-value CDF before correction

p-value CDF after correction
Precision/recall tradeoff

p-value CDF before correction

p-value CDF after correction
Precision/recall tradeoff

$p$-value CDF before correction

$p$-value CDF after correction

$p$-value = 0.05
Sunlight talk

- Overview
- Design
- Evaluation
- Use cases
Sunlight-based tools

Sunlight
(generic, justifiable, and scalable targeting detection)

auditor (e.g., FTC, investigative journalists)
GmailAdObservatory

- Service to **study targeting of Gmail ads** on users’ emails.
  - Meant for researchers and journalists.

- How it works:
  - Researcher supplies a set of emails.
  - GmailAdObservatory uses a set of Gmail accounts to send emails to a separate set of Gmail accounts (the shadows).
  - It then collects ads periodically.
  - Uses Sunlight to detect targeting for each collected ad.
Gmail Targeting Study

- We studied ad targeting in Gmail at pretty large scale.
  - 20K unique ads collected from an inbox with 300 single-keyword emails on various “sensitive” topics.

- Found contradictions to Google’s own privacy statement.

Privacy, Transparency and Choice
[...]
We will also not target ads based on sensitive information, such as race, religion, sexual orientation, health, or sensitive financial categories.

-- http://support.google.com/mail/answer/6603
“We will also not target ads based on sensitive information, such as race, religion, sexual orientation, **health**, or sensitive financial categories.”

<table>
<thead>
<tr>
<th>General Health</th>
<th>email subject &amp; text</th>
<th>ads Title, url &amp; text</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Affordable affordable care [...] (OR)</td>
<td>Illinois Senior Living</td>
<td>p-value = 0.03</td>
</tr>
<tr>
<td></td>
<td>Nursing nursing home [...]</td>
<td><a href="http://www.cottagesofnewlenox.com">www.cottagesofnewlenox.com</a></td>
<td>103 impressions in 36 profiles</td>
</tr>
<tr>
<td></td>
<td>Alzheimer Alzheimer</td>
<td>Assisted Living for Seniors in New Lenox [...]</td>
<td>28% in context</td>
</tr>
<tr>
<td></td>
<td>Alzheimer Alzheimer</td>
<td>1/3 of Seniors 65+ Fall</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depressed depression (OR)</td>
<td>jacuzzi-walk-in-tubs.com/Safety</td>
<td>p-value = 0.01</td>
</tr>
<tr>
<td></td>
<td>Anxious anxious anxiety</td>
<td>Help Eliminate the Fear of Falling in the Bathroom [...]</td>
<td>21 impressions in 8 profiles</td>
</tr>
<tr>
<td></td>
<td>Cancer advice</td>
<td>Is He A Cheater?</td>
<td>p-value = 0.03</td>
</tr>
<tr>
<td></td>
<td>How did you cope with cancer in your family?</td>
<td>spokeo.com/Cheating-Spouse-Search</td>
<td>1179 impressions in 52 profiles</td>
</tr>
<tr>
<td></td>
<td>What an aweful disease!</td>
<td>Enter His Email Address. Find Pics &amp; Profiles From 70+ Social Networks.</td>
<td>20% in context</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The Business of Wellness</td>
<td>p-value = 0.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>healthmediagroup.blogspot.com</td>
<td>380 impressions in 28 profiles</td>
</tr>
<tr>
<td></td>
<td></td>
<td>What my doctor can learn from my Shoe Shine Man [...]</td>
<td>91% in context</td>
</tr>
</tbody>
</table>
“We will also not target ads based on sensitive information, such as race, religion, sexual orientation, health, or **sensitive financial categories.**”

<table>
<thead>
<tr>
<th></th>
<th><strong>email subject &amp; text</strong></th>
<th><strong>ads Title, url &amp; text</strong></th>
<th><strong>Results</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitive Financial</strong></td>
<td><strong>Unemployed</strong> lazy unemployed</td>
<td><strong>Easy Auto Financing</strong>  <a href="http://www.midsouthautoloans.com">www.midsouthautoloans.com</a> Need a quick car loan? We work with credit issues</td>
<td>p-value = 0.006  161 impressions in 24 profiles 8% in context</td>
</tr>
<tr>
<td></td>
<td><strong>Payday</strong> payday loan</td>
<td><strong>Fast Cash Loan Online.</strong>  <a href="http://www.checkintocash.com">www.checkintocash.com</a> Apply Now. Takes Only 5 Minutes. It’s as Easy as 1,2,3.</td>
<td>p-value = 0.007  198 impressions in 10 profiles 6% in context</td>
</tr>
</tbody>
</table>

Notice the extremely low in-context impressions -- the most obscure form of targeting.
$\mathcal{U}DQVSDUHQF\LQIUDVWUXFWXUHIRUWKHGDWDGULYHQZHE$

$\mathcal{V}W$HFWVGDWDXVHIRUWDUJHWLQJDQGSHUVRQDOL\]DWLRQIURP

$\mathcal{V}RQWUROOHGH[\SHULPHQWVZLWKGLIIHUHQWLDWHGSURILOHV$

$\mathcal{V}RUNVDWVFDOHSURYLGHVULJRURXVVWDWLVWLFDOMXVWLILFDWLRQ$

$DQGLVZLGHO\DSSOLFDEOHWRPDQ\VHUYLFHV$

$2XUKRSHLVWRVXSSRUWDQHZJHQHUDWLRQRIWRROVWKDWZLOO$

$NHHSZHEVHUYLFHVPRUHDFFRXQWDEOH$

$\mathcal{V}YHVKRZQWZRWRROVHDFKUHYHDOLQJLQWHUHVWLQJUHVXOWV$

$6XQOLJKWVXPPDU\$
FairTest: fairness testing toolkit for data-driven apps.

[under submission]
Unfair Associations

- Personal data + complex algos can lead to unintended and discriminatory consequences.
- Such consequences are bugs, for which developers should actively test and debug as they do for functionality, performance, reliability bugs.
FairTest

- Testing suite for **unintended associations** in data-driven apps.
  - Detects associations between user attributes (race, gender, age) and service outputs (prices, labels).

- Offers **debugging**, not just detection, capabilities.
Results

- We checked five data-driven apps for unexplained associations, including:
  - Movie recommender.
  - Image labeling system (OverFeat).
  - Predictive healthcare application, the winner of a 2012 Heritage Health Competition.

- We found unexpected associations in all apps, some real bugs.
  - Example: the predictive health app provides good error overall (15%) but its error disproportionately affects elderly patients, where it can be as high as 45%.
My vision for privacy

Critical problem
Erosion of privacy: users share too much, services collect and use their information with almost no accountability.

My vision
Forge a new world where users are privacy aware and services more accountable and privacy-preserving by design.
Related visions

- Two other groups aim to build transparency infrastructures:
  - CMU’s Anupam Datta’s group.
  - Princeton’s Arvind Narayanan and Ed Felten’s group.
  - We uniquely focus on both scalability and broad applicability.

- History:
  - 2014: We published the first paper on this topic: XRay (USENIX Security). Offers good scalability but no statistical justification.
  - 2015: Anupam published AdFisher (PETS). Offers statistical justification but isn’t built to scale with more than one input.
  - 2015: We published Sunlight (CCS). Builds on XRay and AdFisher but offers both scale and statistical justification.