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

	email **subject** & text
E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
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.

ad **title**, url & text	
Ralph Lauren Online Shop www.ralphlauren.com The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1
Cedars Hotel Loughborough www.thecedarshotel.com 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates	Ad2

Example: Gmail Ads

email **subject** **&** text

E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
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.

?

ad **title**, url **&** text

Ralph Lauren Online Shop www.ralphlauren.com The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1
Cedars Hotel Loughborough www.thecedarshotel.com 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates	Ad2

Example: Gmail Ads

email **subject** **&** text

E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
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.

ad **title**, <u>url</u> **&** text

Ralph Lauren Online Shop <u>www.ralphlauren.com</u> The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1
Cedars Hotel Loughborough <u>www.thecedarshotel.com</u> 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates	Ad2

Example: Gmail Ads

	email **subject** & text
E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
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.

?

ad **title**, url & text	
Ralph Lauren Online Shop www.ralphlauren.com The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1
Cedars Hotel Loughborough www.thecedarshotel.com 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates	Ad2

Example: Gmail Ads

email subject & text

E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
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.

ad title, url & text

Ralph Lauren Online Shop www.ralphlauren.com The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1
Cedars Hotel Loughborough www.thecedarshotel.com 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates	Ad2

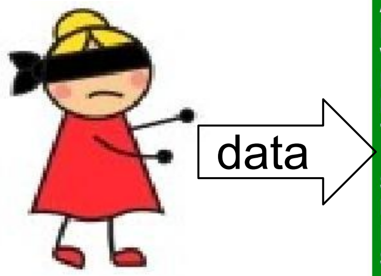
It’s not just Gmail...

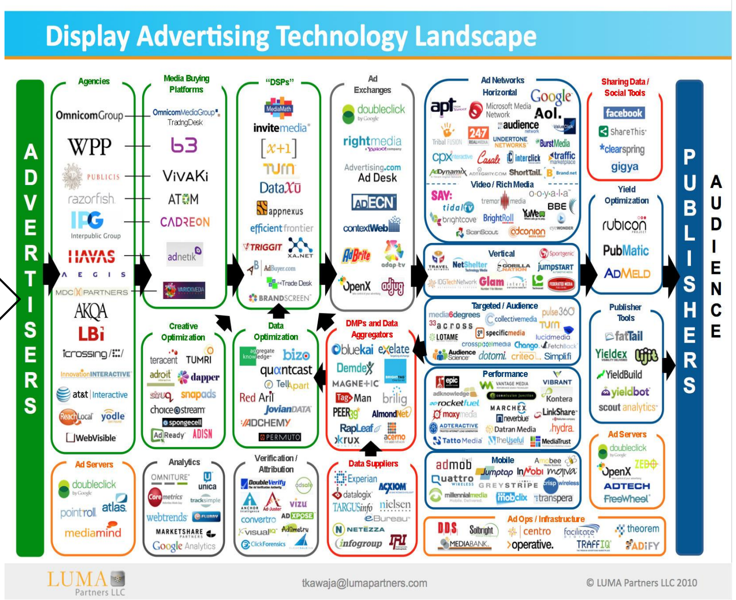
Did you know?

- 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 (eg, 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)

clone me on github

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
E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
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.

ad title, url & text	
Ralph Lauren Online Shop www.ralphlauren.com The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1
Cedars Hotel Loughborough www.thecedarshotel.com 36 Bedrooms, Restaurant, Bar Free WiFi, Parking, Best Rates	Ad2

Example

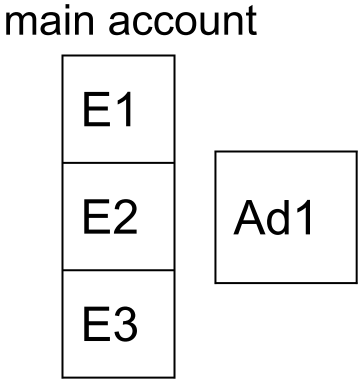
email **subject** **&** text

E1	**Vacation** I'm going on vacation to travel.
E2	**Homosexual** Gay, lesbian, homosexual.
E3	**Pregnant** I'm pregnant. I'm having a baby.

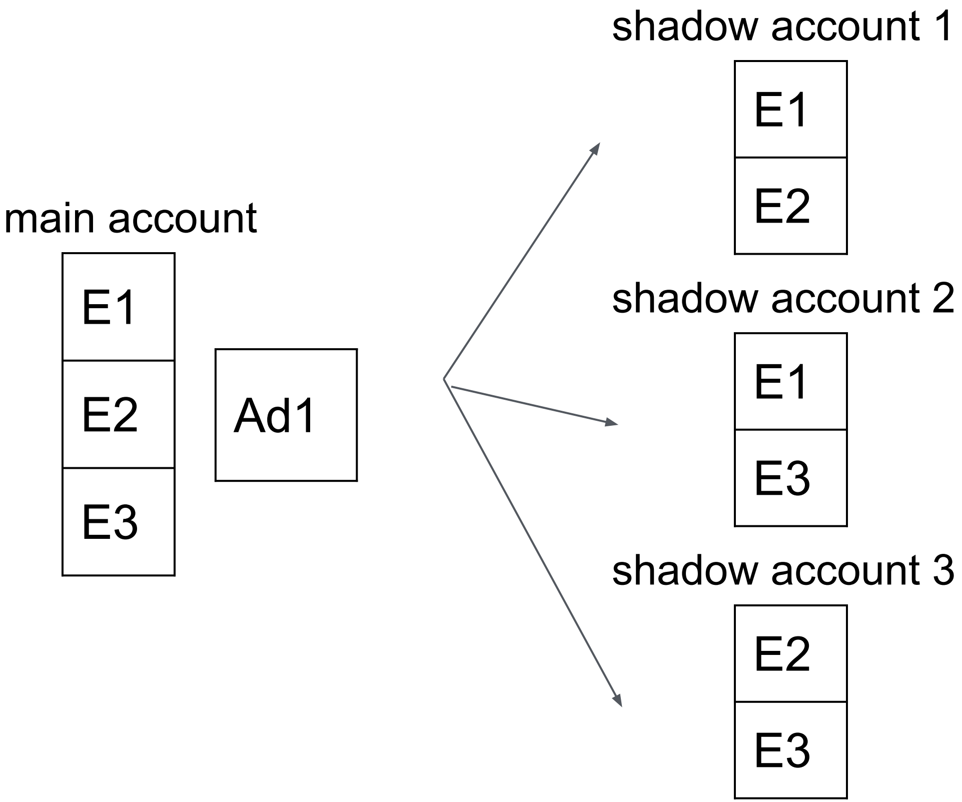
ad **title**, url **&** text

Ralph Lauren Online Shop www.ralphlauren.com The official Site for Ralph Lauren Apparel, Acccessories & More	Ad1

Example

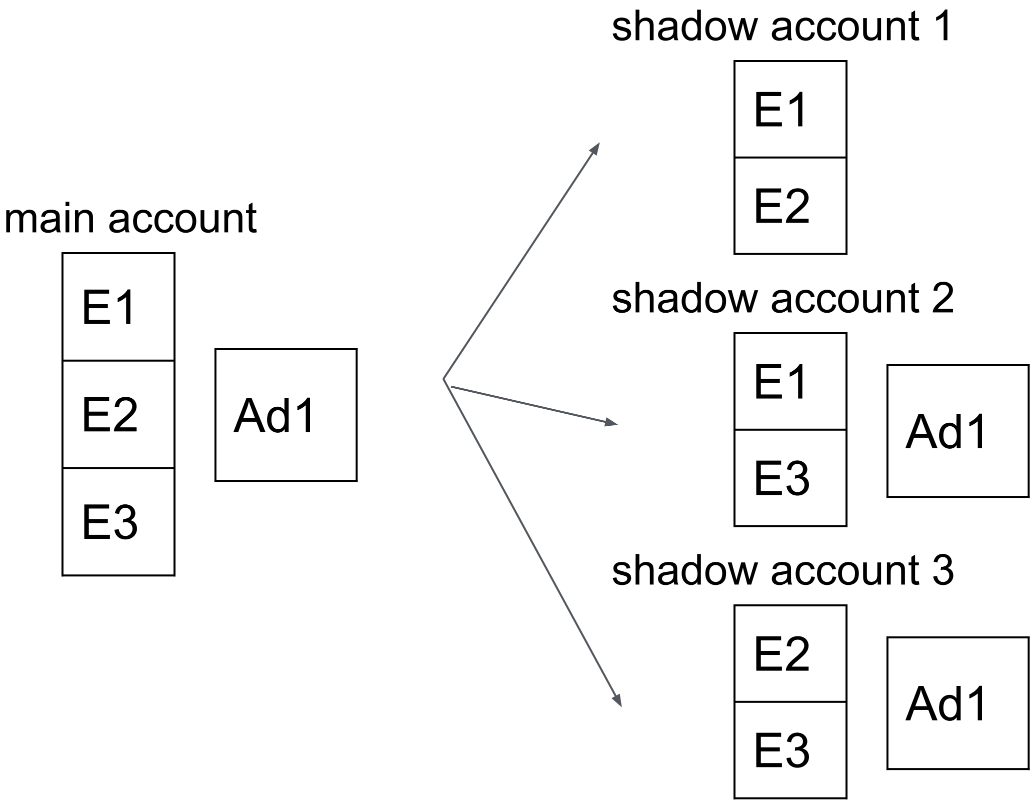


Example

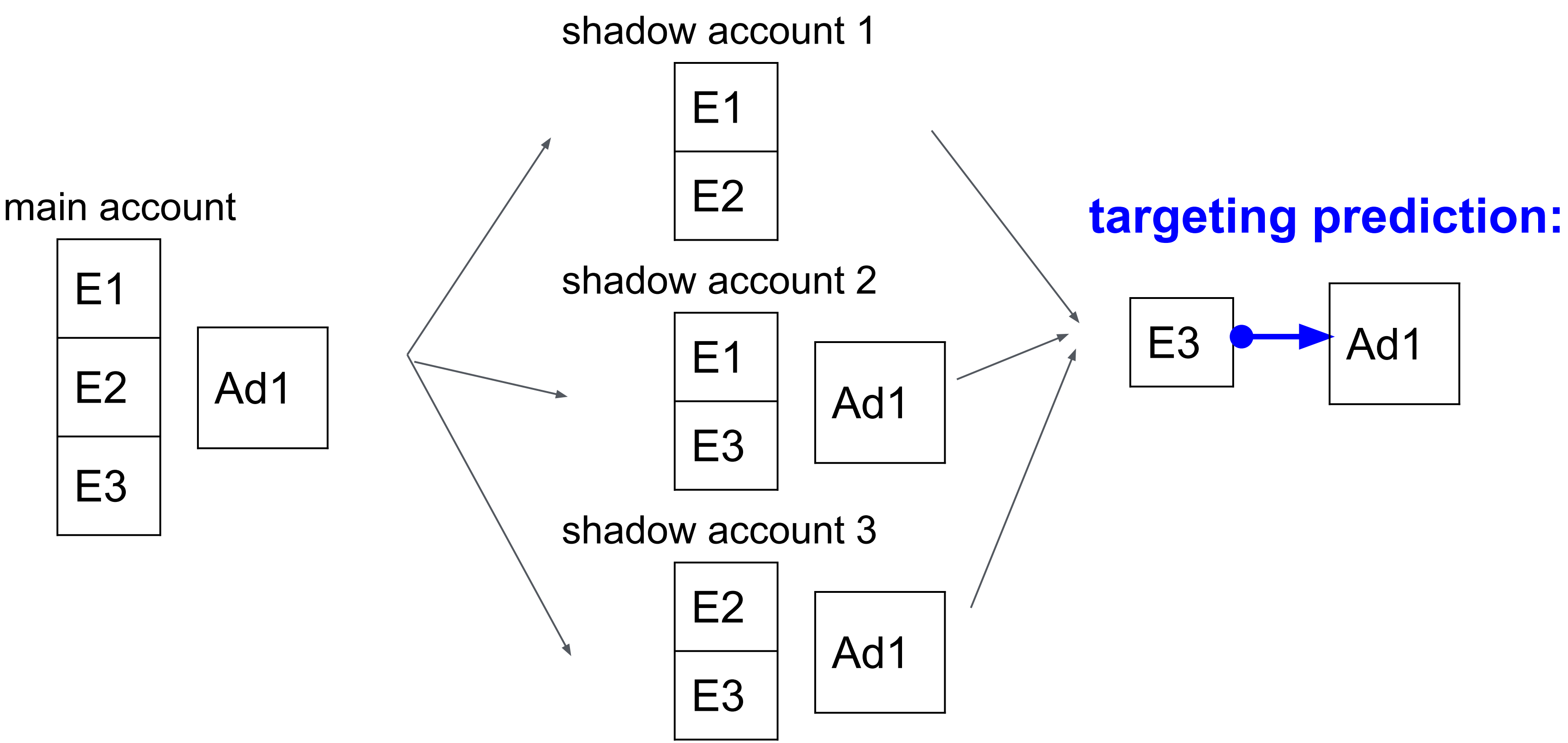


main account
E1
E2
E3
Ad1
shadow account 1
E1
E2
shadow account 2
E1
E3
shadow account 3
E2
E3

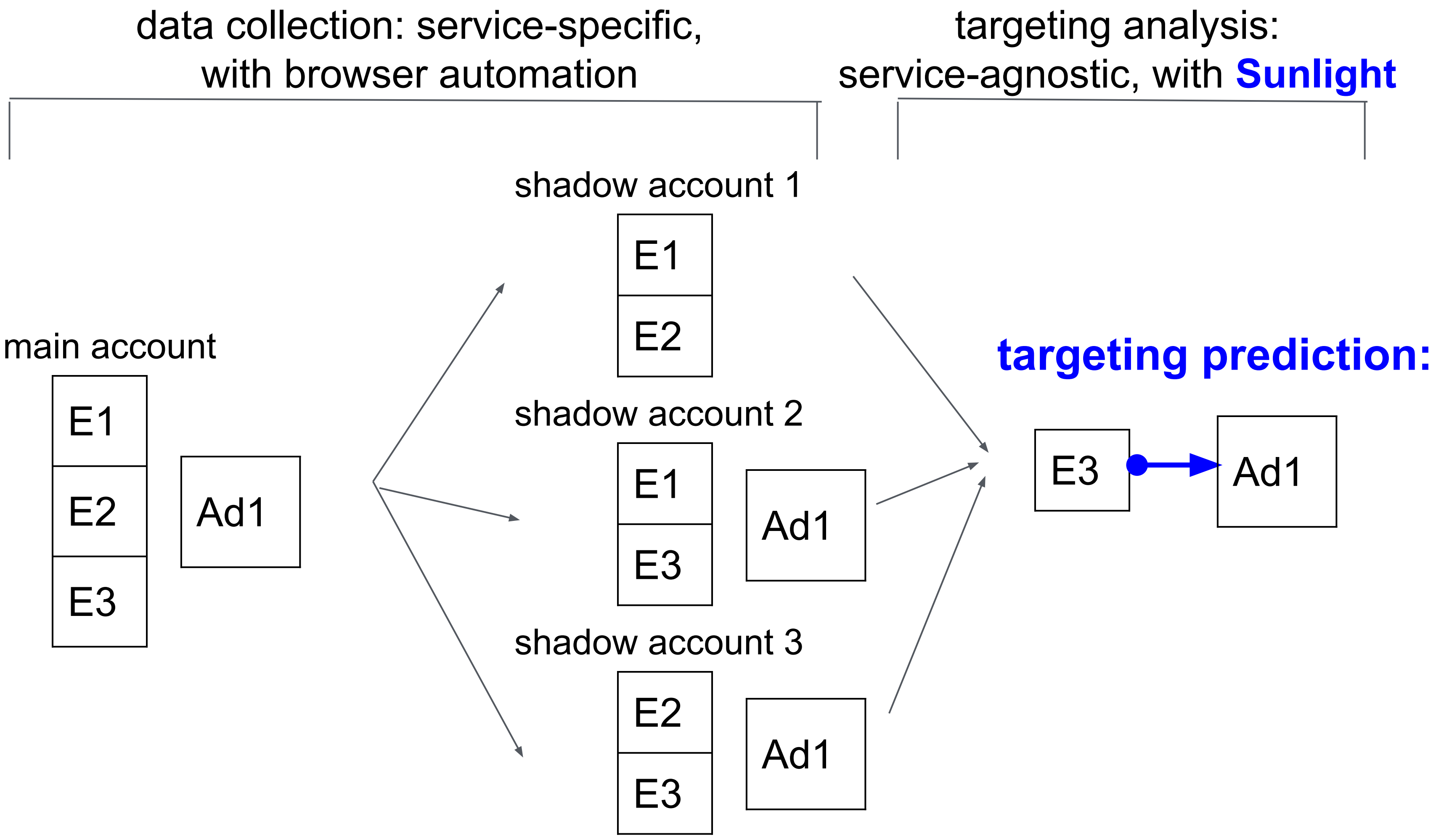
Example


main account
E1
E2
E3
Ad1
shadow account 1
E1
E2
shadow account 2
E1
E3
Ad1
shadow account 3
E2
E3
Ad1

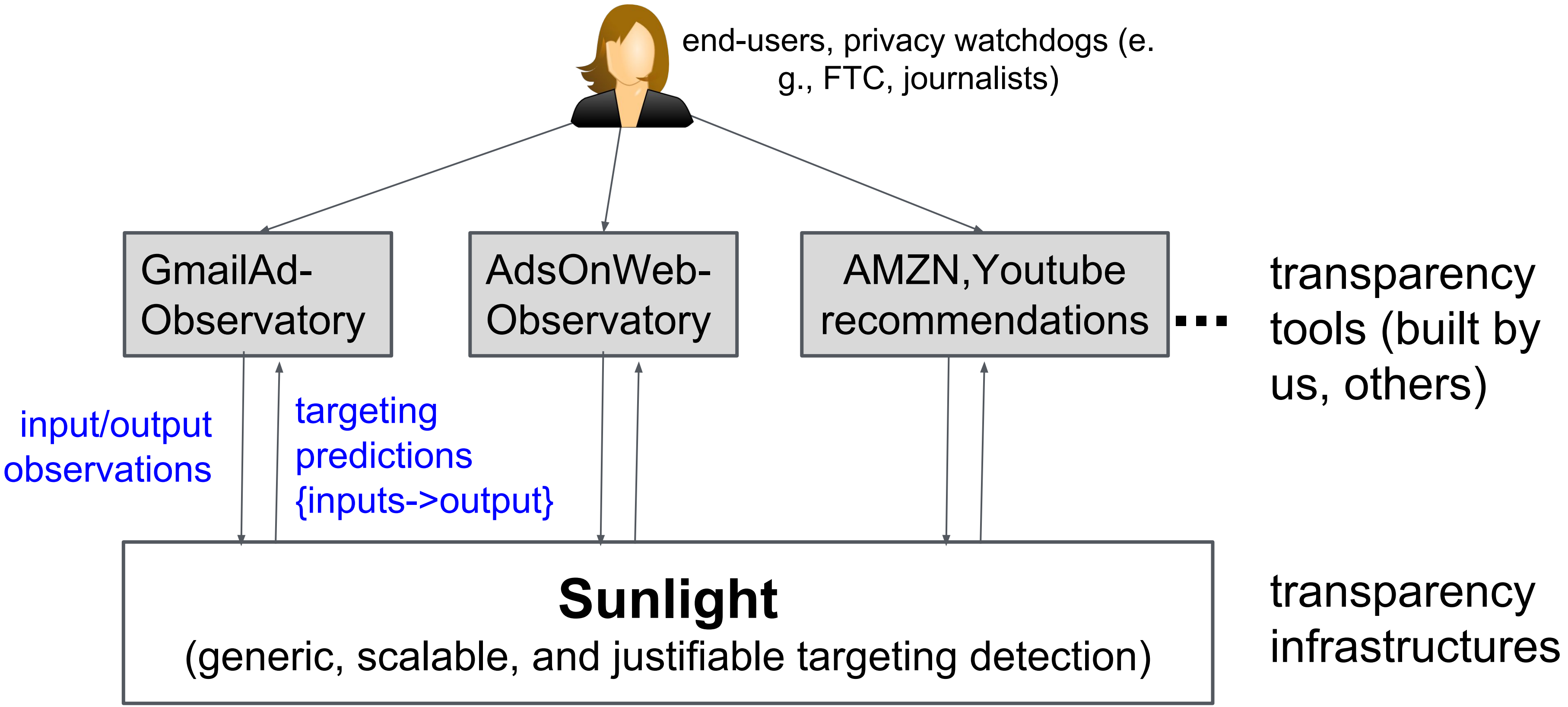
Example



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
targeting analysis:
service-agnostic, with Sunlight
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

Transparency solutions



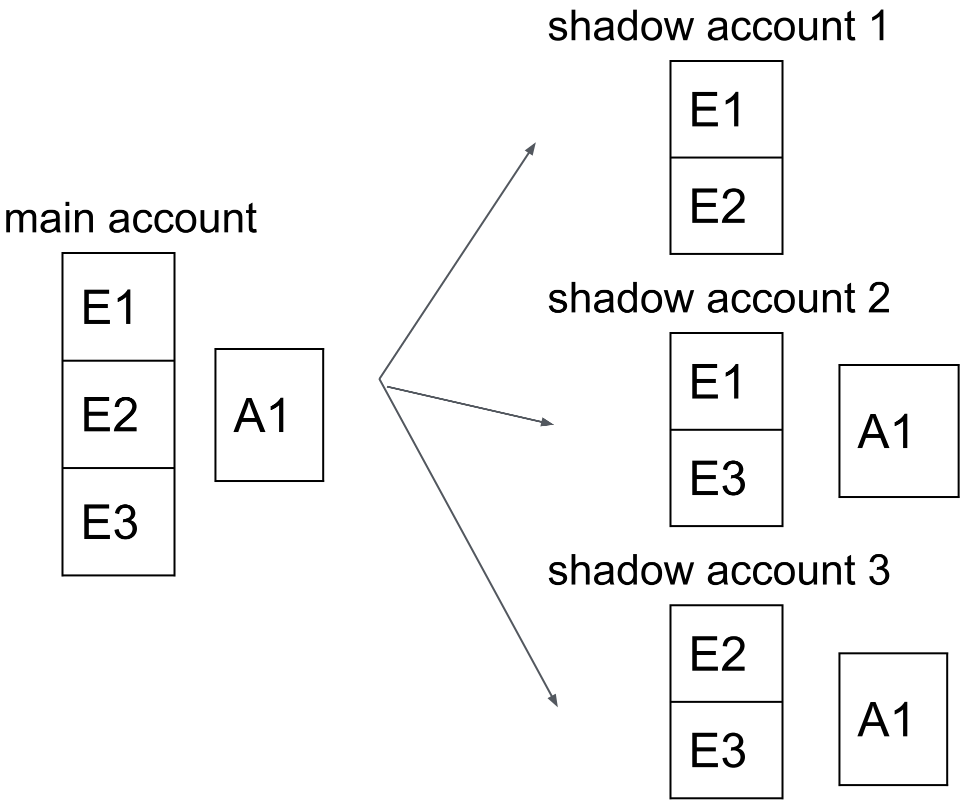
Sunlight talk

- 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 ε > 0 there exists an algorithm that requires C x log(N) accounts to correctly identify the inputs of a targeted output with probability (1 – ε). N is the number of inputs.*

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

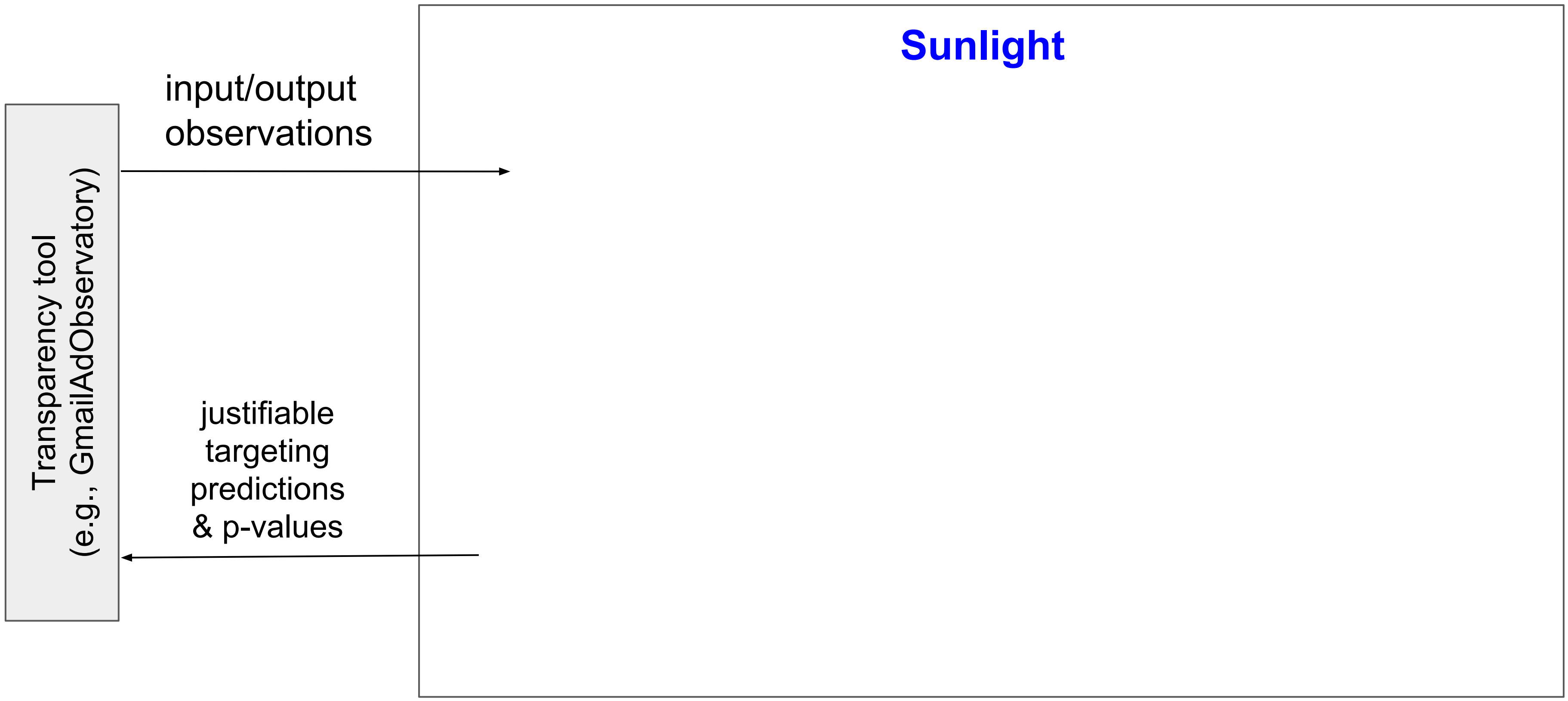
- We incorporate several sparse detection algorithms:
  - **Set intersection** -- simple, not robust
  - **Sparse regressions (Lasso)** -- well established, robust

Justifiable targeting predictions

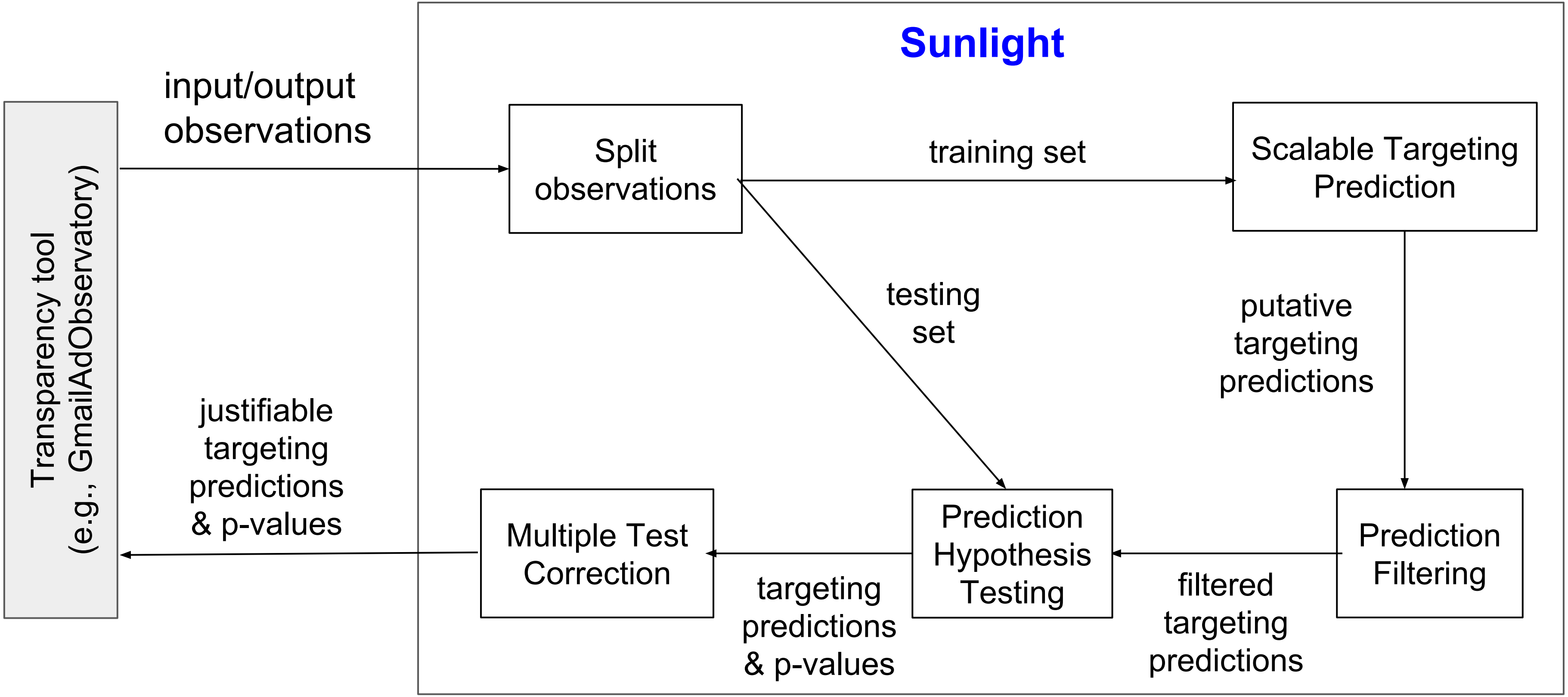
- Sparse algorithms only guarantee asymptotic correctness of the targeting predictions.
- We need correctness assessment for each targeting prediction.

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

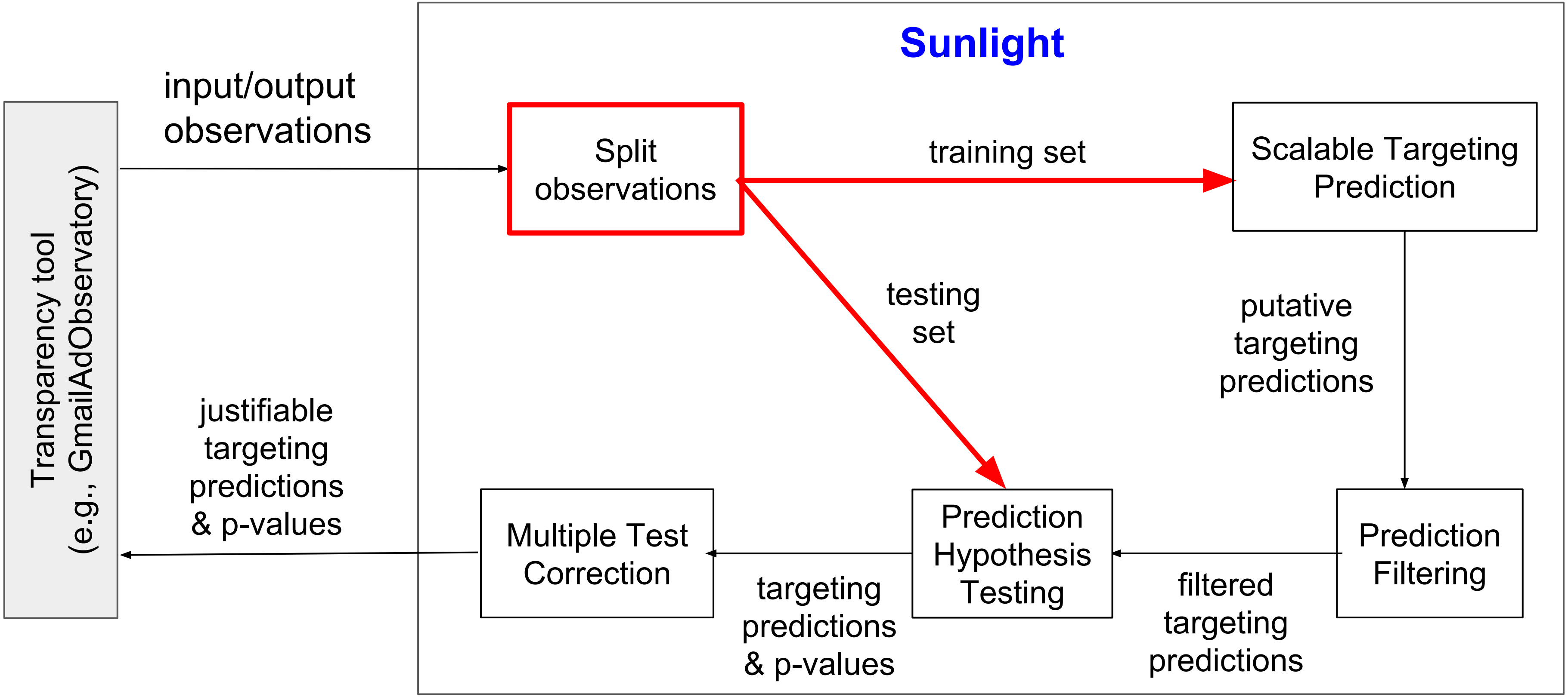
Architecture



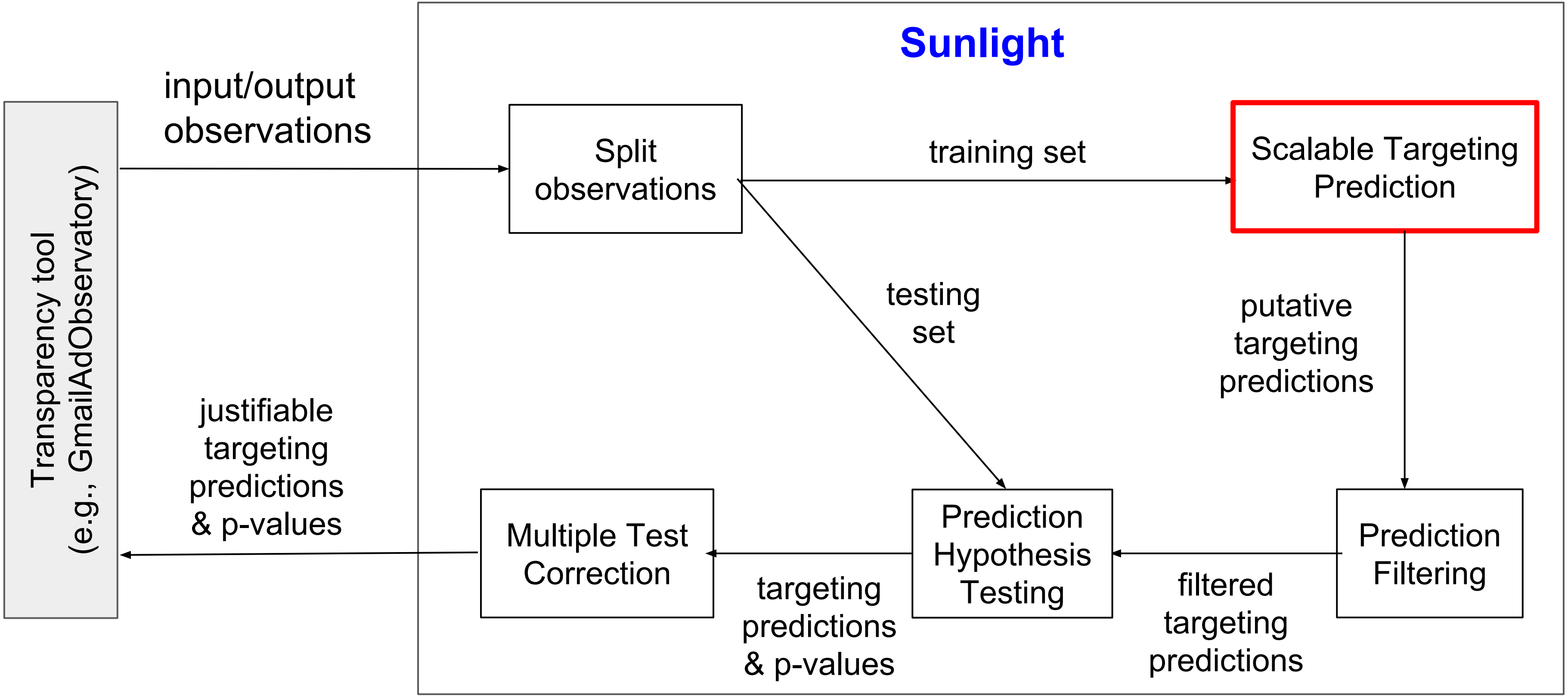
Architecture



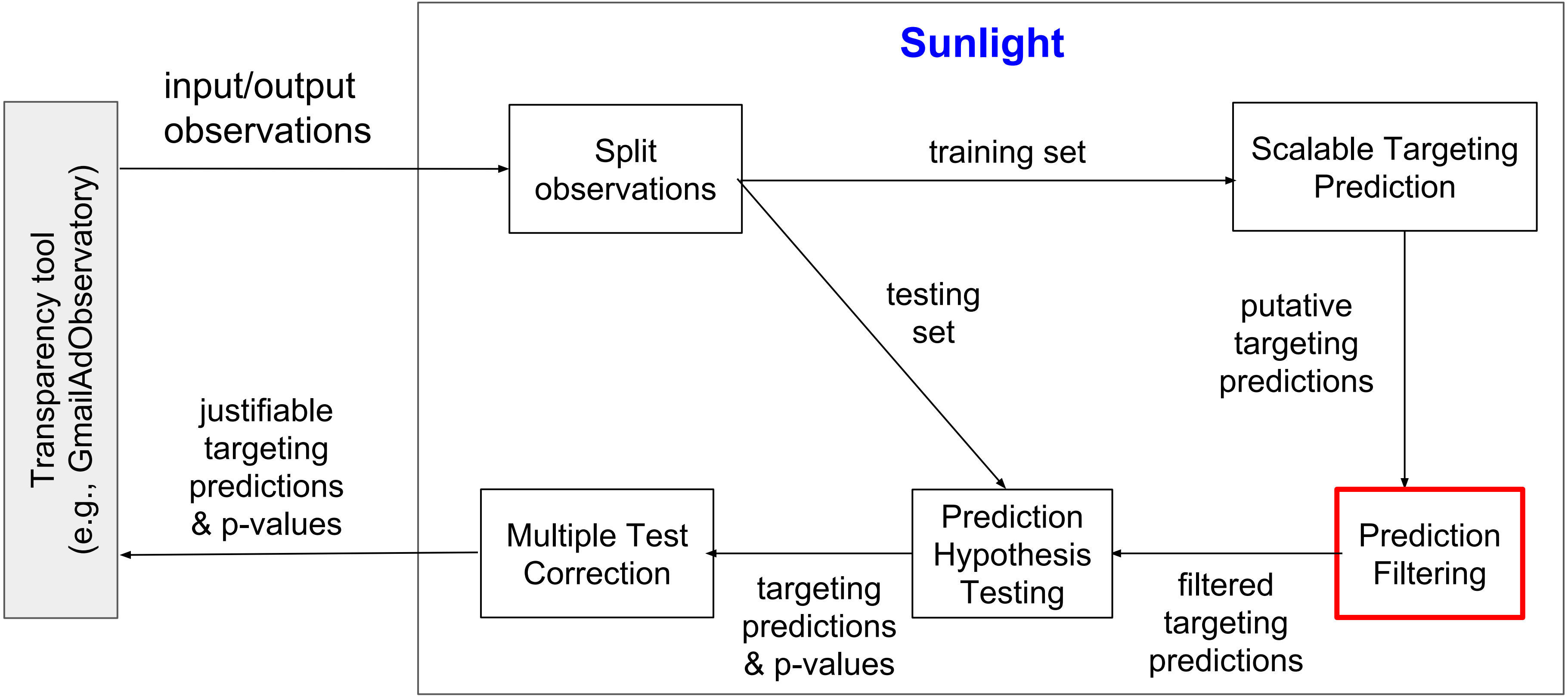
Architecture



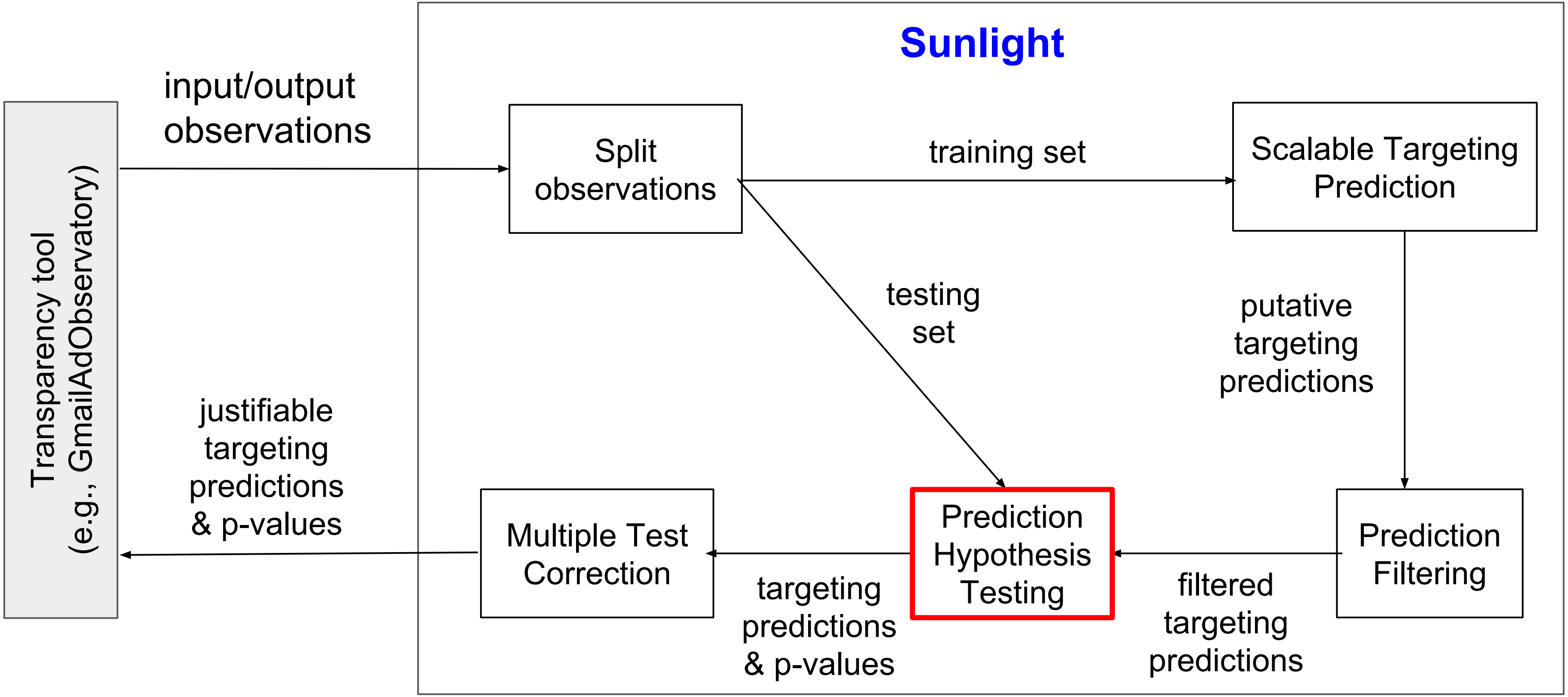
Architecture



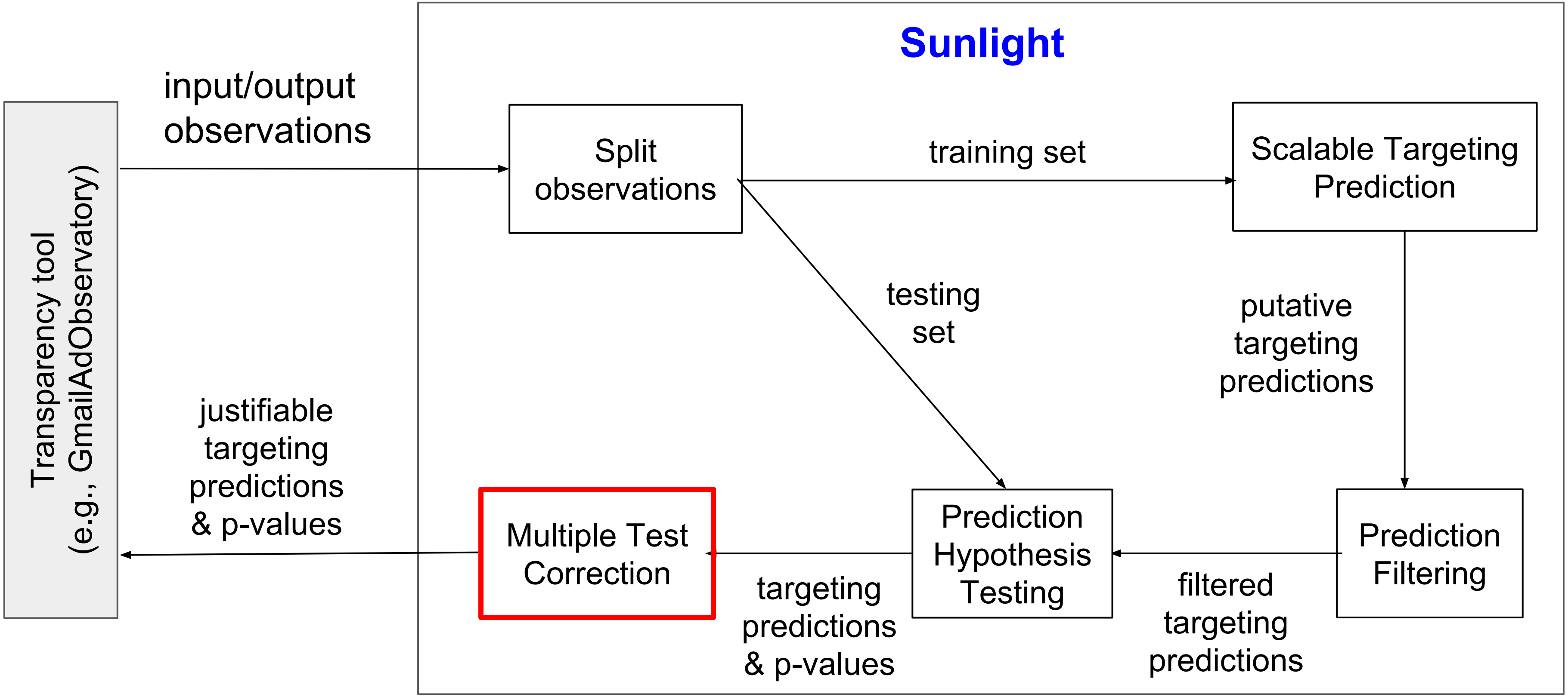
Architecture



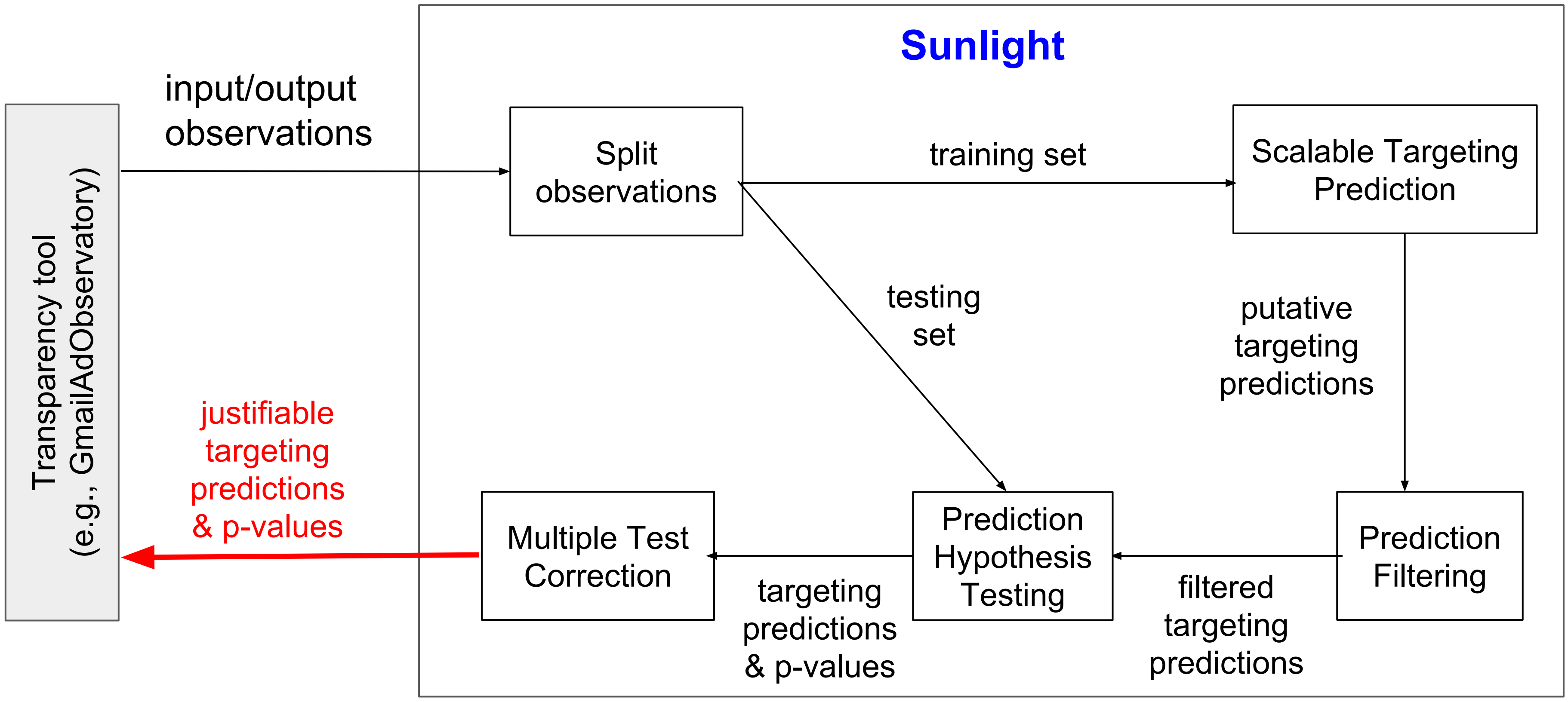
Architecture



Architecture



Architecture



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

Workload	Profiles	Inputs	Outputs
Gmail (one day)	119	327	4099
Website	200	84	4867
Website-large	798	263	19808
YouTube	45	64	308
Amazon	51	61	2593

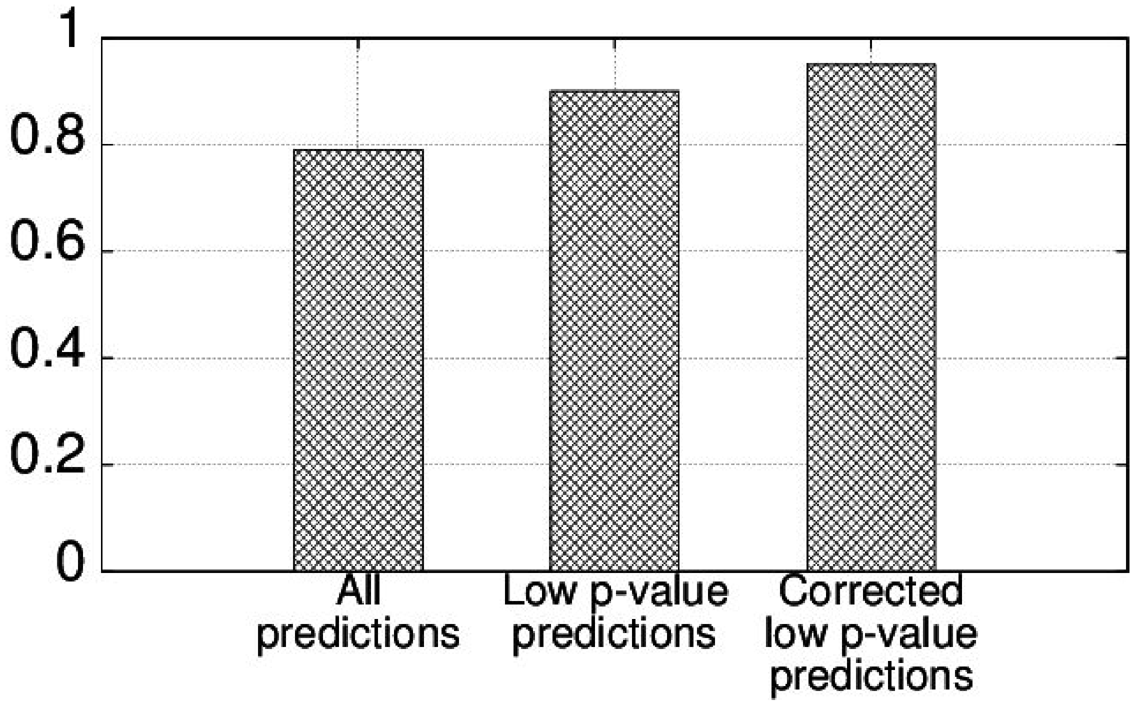
Targeting prediction precision

We developed two methodologies:

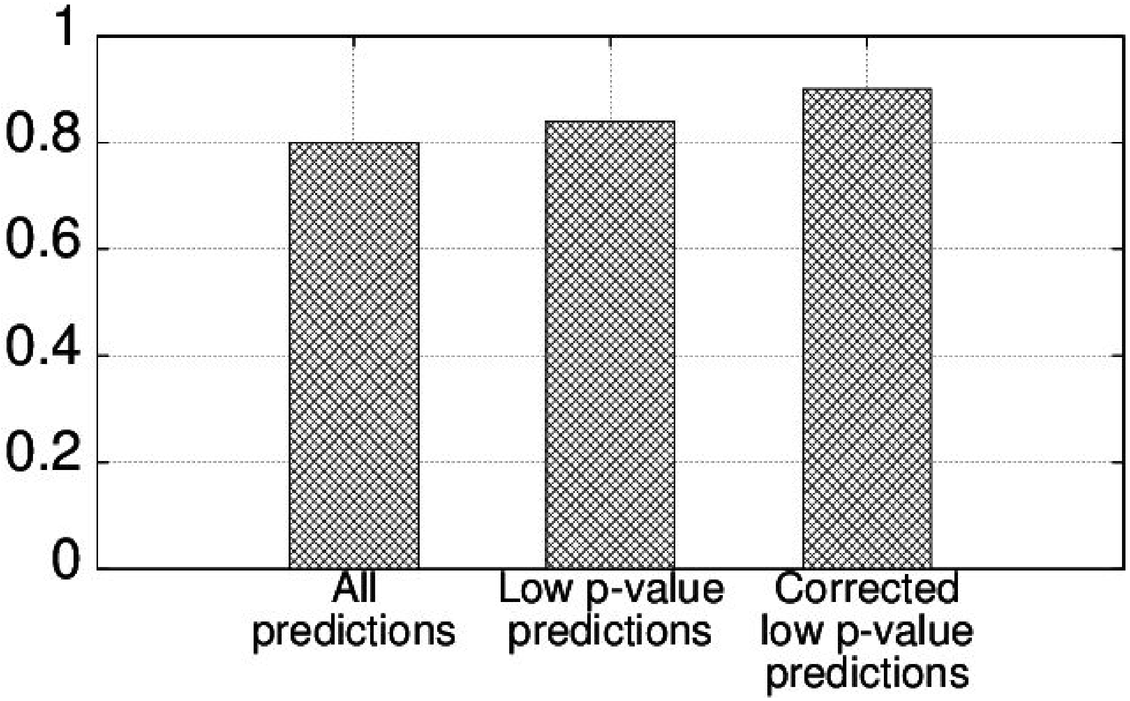
1. Manual assessments of how “believable” are our low-p-value predictions (<0.05).
   - We observed 100% precision for smaller experiments and 95% -96% precision for larger experiments. Despite potential for confirmation bias, this is in line with expectation at p-value < 0.05.
2. Assess the quality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

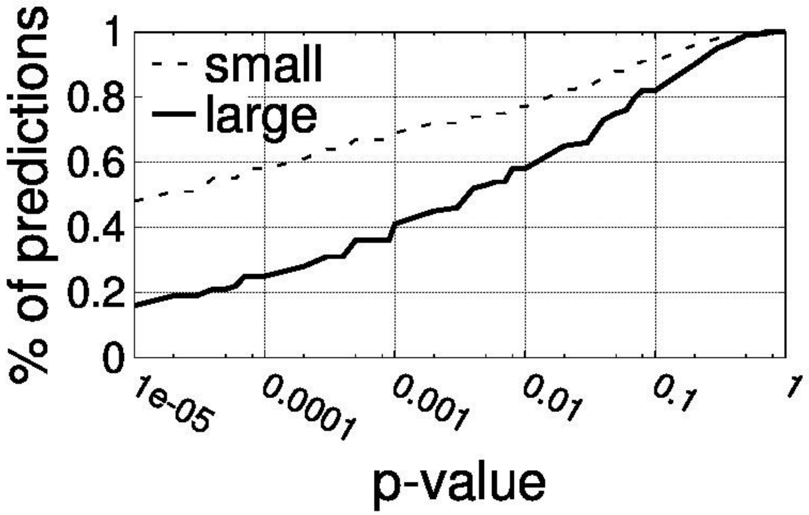


Y: Proportion of success when guessing if an ad will be present in a shadow account.

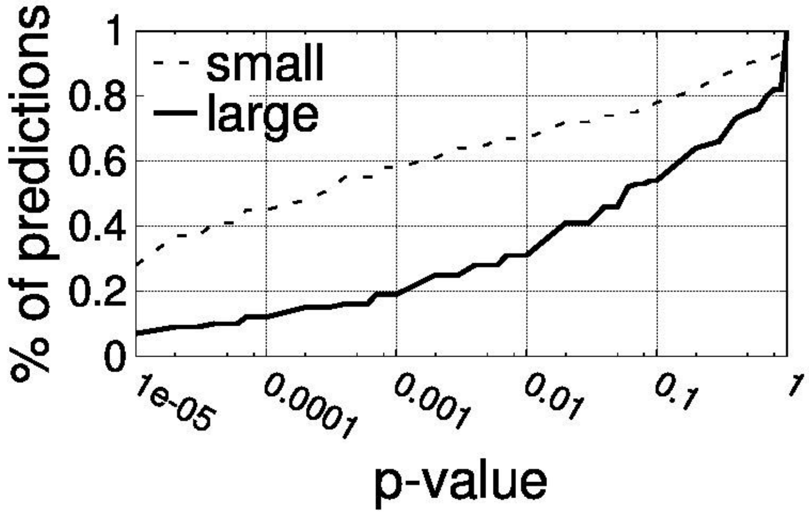
Targeting prediction recall

- 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

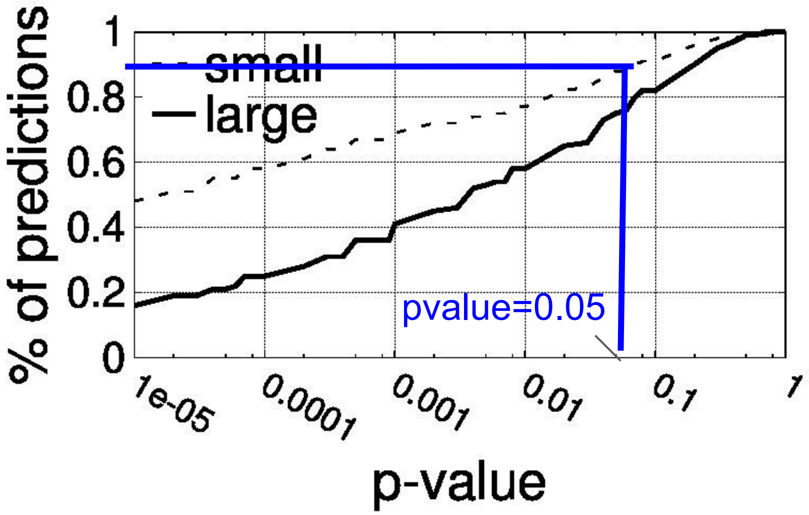


p-value CDF before correction

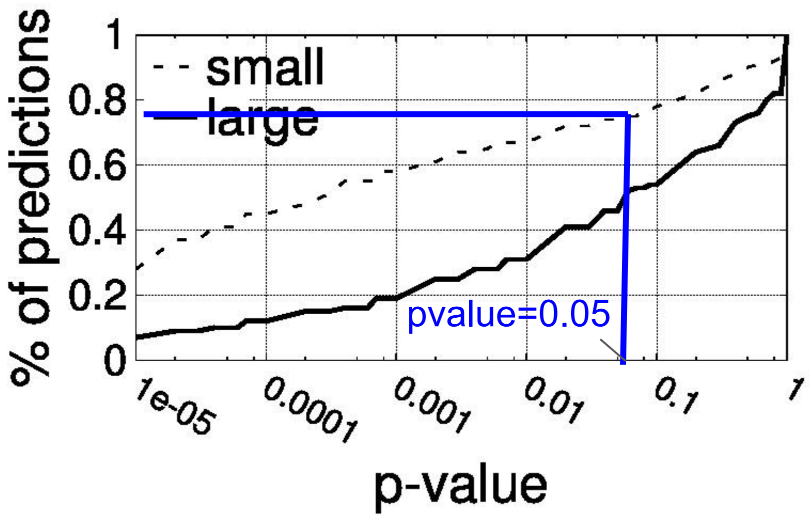


p-value CDF after correction

Precision/recall tradeoff

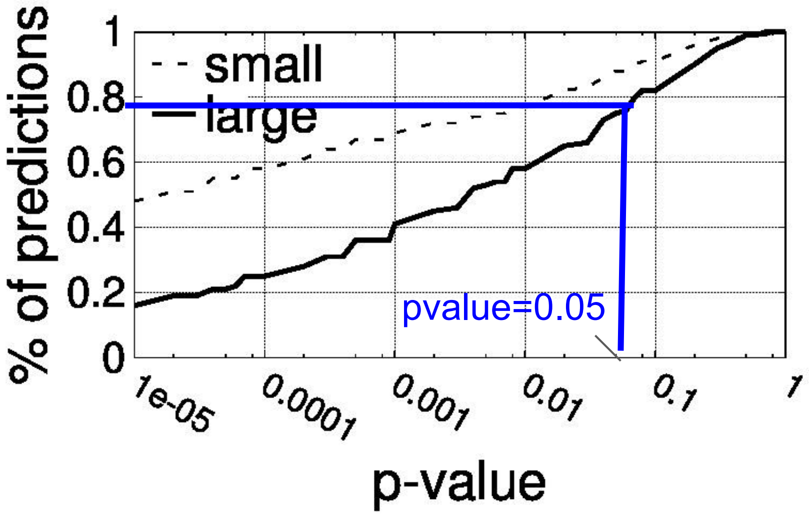


p-value CDF **before correction**

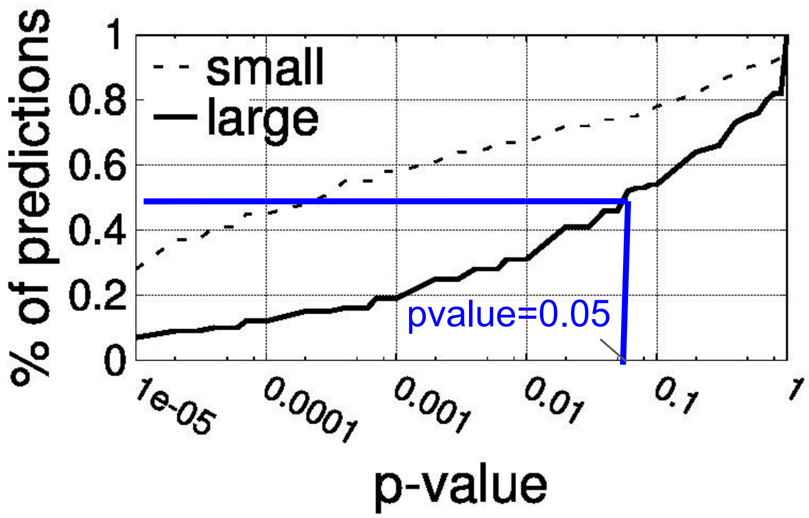


p-value CDF **after correction**

Precision/recall tradeoff



p-value CDF before correction



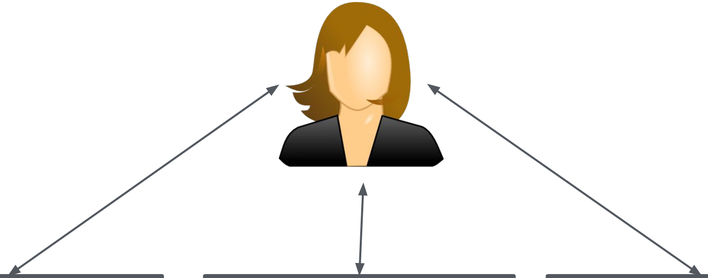
p-value CDF after correction

Sunlight talk

- Overview
- Design
- Evaluation
- Use cases

Sunlight-based tools

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



GmailAd-Observatory	AdsOnWeb-Observatory	Data-Observatory	Amazon, Youtube recommendations

Sunlight

(generic, justifiable, and scalable targeting detection)

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.”

	email subject & text	ads Title, url & text	Results
General Health	**Affordable** afforable care [...] **(OR)** **Nursing** nursing home [...]	**Illinois Senior Living** www.cottagesofnewlenox.com Assisted Living for Seniors in New Lenox [...]	p-value = 0.03 103 impressions in 36 profiles 28% in context
	Alzheimer Alzheimer Alzheimer	**1/3 of Seniors 65+ Fall** jacuzzi-walk-in-tubs.com/Safety Help Eliminate the Fear of Falling in the Bathroom [...]	p-value = 0.01 21 impressions in 8 profiles 100% in context
	Depressed depression **(OR)** **Anxious** anxious anxiety	**Is He A Cheater?** spokeo.com/Cheating-Spouse-Search Enter His Email Address. Find Pics & Profiles From 70+ Social Networks.	p-value = 0.03 1179 impressions in 52 profiles 20% in context
	Cancer advice How did you cope with cancer in your familly? What an aweful disease!	**The Business of Wellness** healthmediagroup.blogspot.com What my doctor can learn from my Shoe Shine Man [...]	p-value = 0.04 380 impressions in 28 profiles 91% in context

“We will also not target ads based on sensitive information, such as race, religion, sexual orientation, health, or **sensitive financial categories**.”

	email subject & text	ads Title, url & text	Results
Sensitive Financial	**Unemployed** lazy unemployed	**Easy Auto Financing** www.midsouthautoloans.com Need a quick car loan? We work with credit issues	p-value = 0.006 161 impressions in 24 profiles 8% in context
	Payday payday loan	**Fast Cash Loan Online.** www.checkintocash.com Apply Now. Takes Only 5 Minutes. It's as Easy as 1,2,3.	p-value = 0.007 198 impressions in 10 profiles 6% in context

Notice the extremely low in-context impressions -- the most obscure form of targeting.

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.

THE WALL STREET JOURNAL.

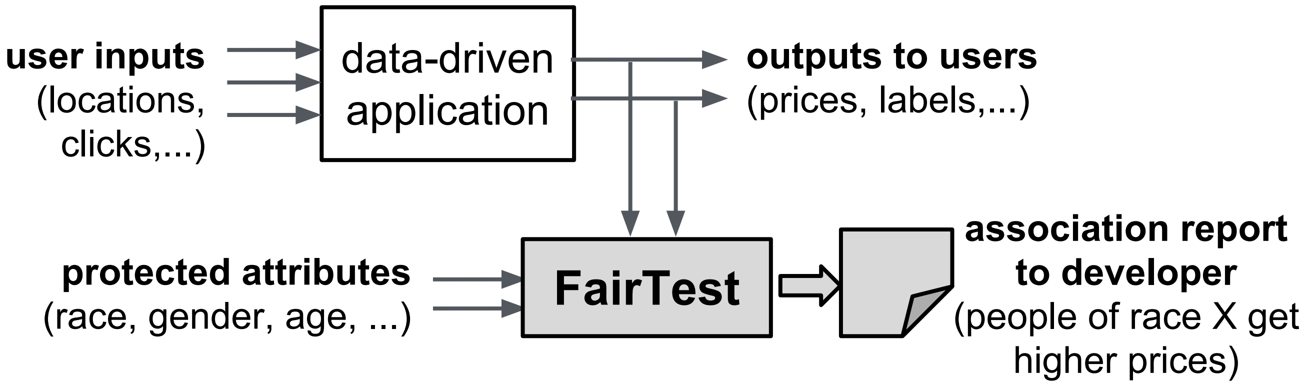
In what appears to be an unintended side effect of Staples' pricing methods—areas that tended to see the discounted prices had a higher average income than areas that tended to see higher prices.

c|net Search CNET Reviews News

Google apologizes for algorithm mistakenly calling black people 'gorillas'

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.