UNIVERSITÀ
DEGLI STUDI
DI PADOVA

PPIA: PEEPING THROUGH THE OPEN DOOR

PRIVACY PRESERVING INFORMATION ACCESS

PhD in Information Engineering
A.Y. 2025/2026

GUGLIELMO FAGGIOLI
Intelligent Interactive Information Access (IIIA) Hub
Department of Information Engineering
University of Padua

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

MYSELF

I am Guglielmo Faggioli, Post-doc researcher at the University of Padova.

My main research topics are:

- Privacy preserving information access
  - Query obfuscation for IR
  - Data anonymization
- Information retrieval
  - Evaluation
  - Performance prediction
  - Dense models

WHAT IS PRIVACY

Defining privacy is by itself a challenging task - it is something you have until you don’t.

We are all aware of what the absence of privacy causes:

- the uneasing feeling of being spied upon;
- the unpleasant sensation caused by someone knowing “too much”;
- the urge of behaving in a certain way because “someone is watching us”.

WHAT IS PRIVACY

Admit it … it happened to everybody:

“I’m sure my phone is spying me!!! I have only talked about <insert something you talked about here>, and now I have plenty of ads about it!!!”

What effect does this have on the trust and privacy perception that you have about a service?

WHAT IS PRIVACY

These are **pure conspiracy theories**, but the **unpleasant feeling** is very real!

What if our behaviour and design choices as computer engineers induce the same feeling on our users?

WHAT IS PRIVACY

We are **genetically encoded to care about privacy** - our (very old) ancestors hated being watched upon because usually the one watching was a predator!

Nevertheless privacy is also linked to more practical needs.

Even in our modern society, several perfectly legal **behaviours** are considered not **socially acceptable**: we do not want to draw **social stigma** upon us.

WHAT IS PRIVACY

For those of you who prefer a more formal definition:

“someone's right to keep their personal matters and relationships secret”

Cit. Cambridge Dictionary

WHAT IS PRIVACY

For those of you who prefer a more formal definition:

“someone's right to keep their personal matters and relationships secret”

it has something to do with the law: it is a right

Cit. Cambridge Dictionary

WHAT IS PRIVACY

For those of you who prefer a more formal definition:

“someone's right to keep their personal matters and relationships secret”

it regards the personal sphere of interest

Cit. Cambridge Dictionary

WHAT IS PRIVACY

For those of you who prefer a more formal definition:

“someone's right to keep their personal matters and relationships secret”

it implies secrecy

Cit. Cambridge Dictionary

WHAT IS PRIVACY

- Should a court be able to investigate on private documents, if they *might* contain evidence of a felony?

- Should a journalist expose a politician, which is deeply involved in writing a prohibitionist law, that is using drugs?

- Should your neighbor peep through your open door?

WHAT VALUE DO HUMANS GIVE TO PRIVACY?

“No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.”

Universal Declaration of Human Rights, Art. 12

WHAT VALUE DO HUMANS GIVE TO PRIVACY?

We consider the **right to privacy as innate in the human being** - it should be our moral duty to protect it!

I think there is far greater value in a top performing algorithm than in ethical duty…

WHAT VALUE DOES THE SOCIETY GIVE TO PRIVACY?

Privacy is strictly regulated legally.

- National laws
- International frameworks
- Legal boards

Some examples are Health Insurance Portability and Accountability Act (**HIPAA**), General Data Protection Regulation (**GDPR**), European Data Protection Board (**EDPB**).

WHAT VALUE DOES THE SOCIETY GIVE TO PRIVACY?

Violating the GDPR might cost you **4% of your revenue or €20 millions** (depending on which one is the **highest**!).

Legally, you are obliged to preserve your user’s privacy!

but my service is the best in town,
plus if nobody knows…

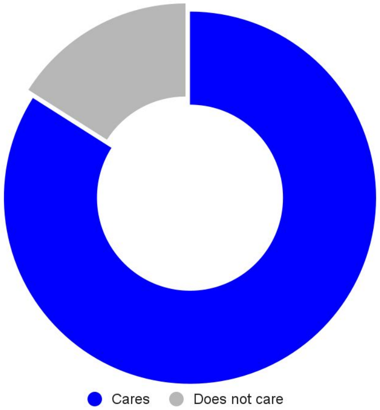
WHAT VALUE DO USERS GIVE TO PRIVACY?

USERS CARE ABOUT PRIVACY. A LOT.

WHAT VALUE DO USERS GIVE TO PRIVACY?

84% of the users **care** about their privacy:

- I care about data privacy
- I care about protecting others
- I want more control

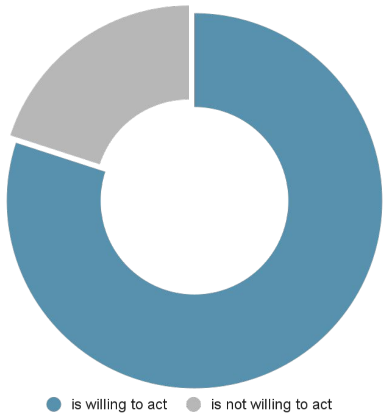


CISCO consumer privacy survey

WHAT VALUE DO USERS GIVE TO PRIVACY?

80% of the users **are willing to act**:

- I am willing to spend time and money to protect my data
- This is a buying factor for me
- I expect to pay more

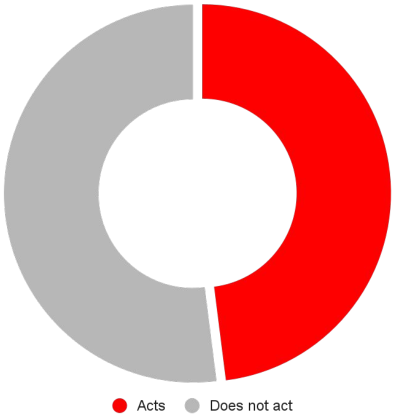


CISCO consumer privacy survey

WHAT VALUE DO USERS GIVE TO PRIVACY?

48% users **act**:

- I have switched companies or providers over their data policies or data sharing practices



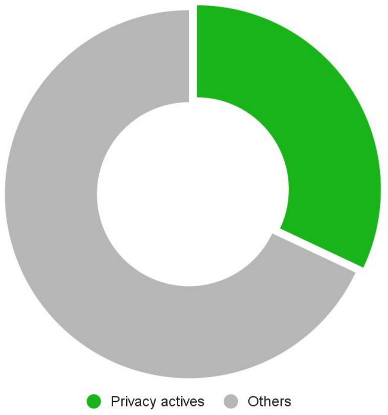
If you don’t have good privacy policies **HALF** of the users will leave your service!!!!

CISCO consumer privacy survey

WHAT VALUE DO USERS GIVE TO PRIVACY?

Combining all these characteristics, we obtain that 32% of the users are **Privacy Actives: they care and value their privacy, and will be ready to leave your service if they do not feel protected.**

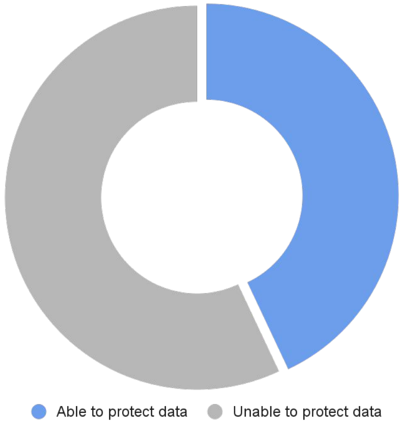
Others, even though won't leave immediately, will have low trust your service.

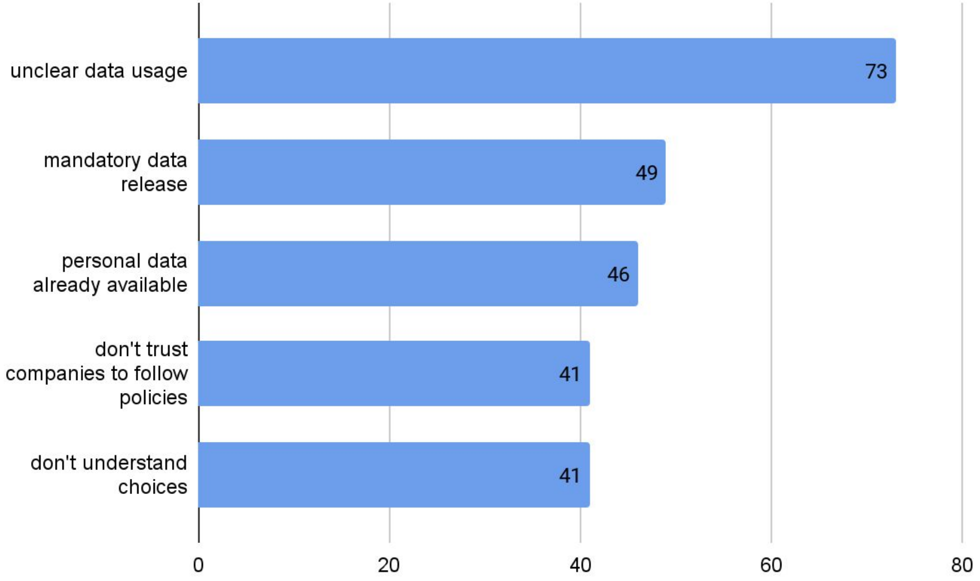


CISCO consumer privacy survey

WHAT VALUE DO USERS GIVE TO PRIVACY?

43% of the users feel unable to protect their data





CISCO consumer privacy survey

WHAT VALUE DO USERS GIVE TO PRIVACY?

If the ethics nor the legal aspects of the matter concern you too much, be aware that, by ignoring privacy, **your company might encounter huge issues in terms of revenues.**

WHAT VALUE DO **YOU** GIVE TO YOUR PRIVACY?

GENERAL DATA PROTECTION REGULATION

GDPR: PRINCIPLES

Data should be handled with the following principles in mind (Article 5):

- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimization
- Accuracy
- Storage Limitation
- Integrity and confidentiality

PROTECTION BY DESIGN AND BY DEFAULT

The system **needs to be designed** to protect the users’ privacy - it is mandatory to already think (and be able to prove) that privacy concerns have been addressed at design time.

The system needs to protect the privacy **by default**: you don’t have to ask the user to specify if they want their privacy to be protected.

GDPR: DATA PROCESSING

You can process data only in the following cases:

- Life or death matter
- Public interest
- you have a legal obligation to do so
- you are about to enter in contract with the subject
- specific consent from the subject was given
- you have legitimate interest

GDPR: CONSENT

Data usage in the GDPR framework revolves around the consent:

- Freely given, specific, informed and unambiguous
- Clear
- Can be withdrawn at any moment
- You need to have documentary evidence

GDPR: RIGHTS

Right to be informed (Article 12, Article 13, and Article 14)

Right of access (Article 15)

Right to data portability (Article 20)

Right to rectification (Article 16)

Right to erasure (Article 17)

Right to restriction of processing (Article 18)

Right to object (Article 21)

to automated decision making and profiling (Article 22)

Recital 26

The principles of data protection should apply to any information concerning an identified or identifiable natural person.

Personal data which have undergone pseudonymisation, [...] should be considered to be information on an identifiable natural person.

To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used [...].

[...] account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

The principles of data protection should therefore not apply to anonymous information [...]. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

INFORMATION ACCESS

A CAVEAT BEFORE CONTINUING

Protecting users privacy means working on three separate aspects:

- Avoid unauthorized, fraudulent and mischievous accesses to the data
- Ensuring proper usage of the data
- Make sure that no other information besides the one actually given is available

A CAVEAT BEFORE CONTINUING

Protecting users privacy means working on three separate aspects:

- Avoid unauthorized, fraudulent and mischievous accesses to the data;
- Ensuring proper usage of the data;
- Make sure that no other information besides the one actually given is available.

A CAVEAT BEFORE CONTINUING

Protecting users privacy means working on three separate aspects:

- Avoid unauthorized, fraudulent and mischievous accesses to the data;
- Ensuring proper usage of the data;
- Make sure that no other information besides the one actually given is available.

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Recommender systems should be the safest kind of information access systems in terms of privacy:

- we give some information to the system (likes, views, thumbs up, stars, interactions, purchases, etc.)

- we receive improved recommendations

No one besides the algorithm should and could see our data …

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

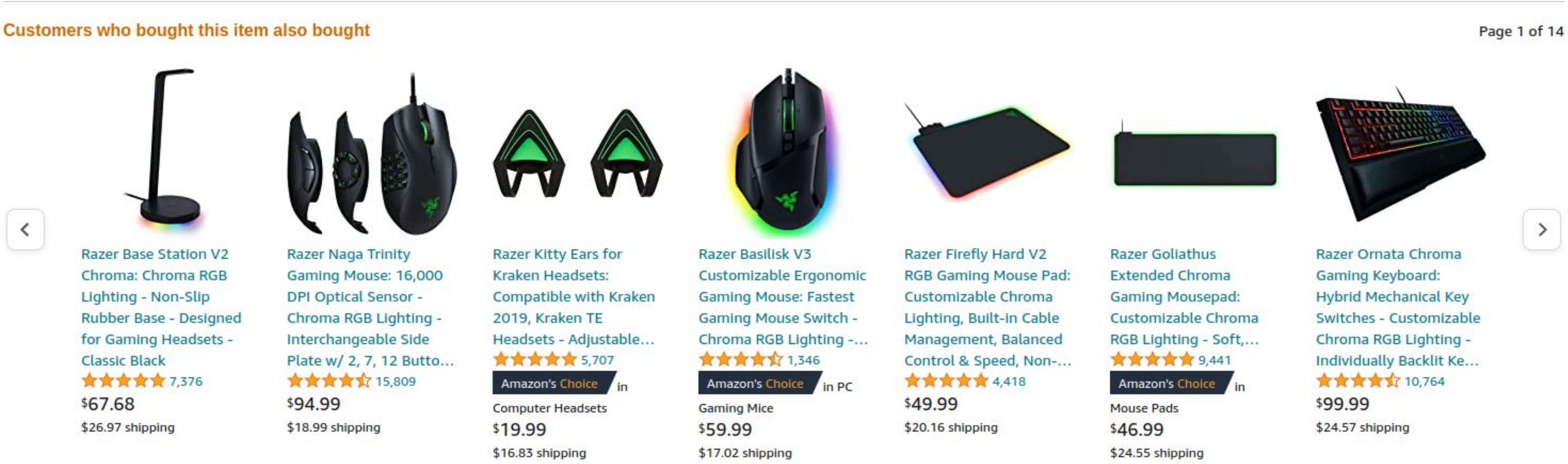
Recommender systems should be the safest kind of information access systems in terms of privacy:

- we give some information to the system (likes, views, thumbs up, stars, interactions, etc.)

- we receive improved recommendations

No one besides the algorithm should and could see our data … or do they?

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS



Customers who bought this item also bought

Page 1 of 14

Razer Base Station V2 Chroma: Chroma RGB Lighting - Non-Slip Rubber Base - Designed for Gaming Headsets - Classic Black
7,376
$67.68
$26.97 shipping

Razer Naga Trinity Gaming Mouse: 16,000 DPI Optical Sensor - Chroma RGB Lighting - Interchangeable Side Plate w/ 2, 7, 12 Butto...
15,809
$94.99
$18.99 shipping

Razer Kitty Ears for Kraken Headsets: Compatible with Kraken 2019, Kraken TE Headsets - Adjustable...
5,707
Amazon's Choice in Computer Headsets
$19.99
$16.83 shipping

Razer Basilisk V3 Customizable Ergonomic Gaming Mouse: Fastest Gaming Mouse Switch - Chroma RGB Lighting -...
1,346
Amazon's Choice in PC Gaming Mice
$59.99
$17.02 shipping

Razer Firefly Hard V2 RGB Gaming Mouse Pad: Customizable Chroma Lighting, Built-in Cable Management, Balanced Control & Speed, Non-...
4,418
$49.99
$20.16 shipping

Razer Goliathus Extended Chroma Gaming Mousepad: Customizable Chroma RGB Lighting - Soft,...
9,441
Amazon's Choice in Mouse Pads
$46.99
$24.55 shipping

Razer Ornata Chroma Gaming Keyboard: Hybrid Mechanical Key Switches - Customizable Chroma RGB Lighting - Individually Backlit Ke...
10,764
$99.99
$24.57 shipping

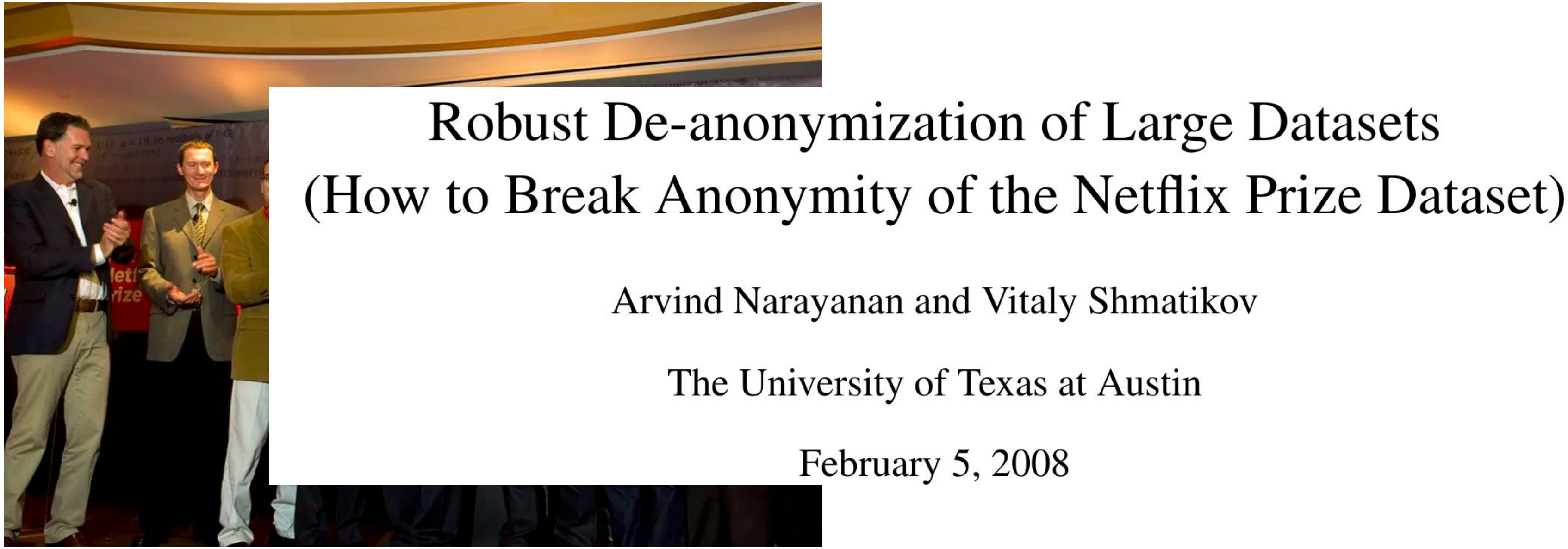
For a keyboard is fine, but what about **shares**? do we want our marketing information to be used by someone else?

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Still…

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Still…



Robust De-anonymization of Large Datasets
(How to Break Anonymity of the Netflix Prize Dataset)

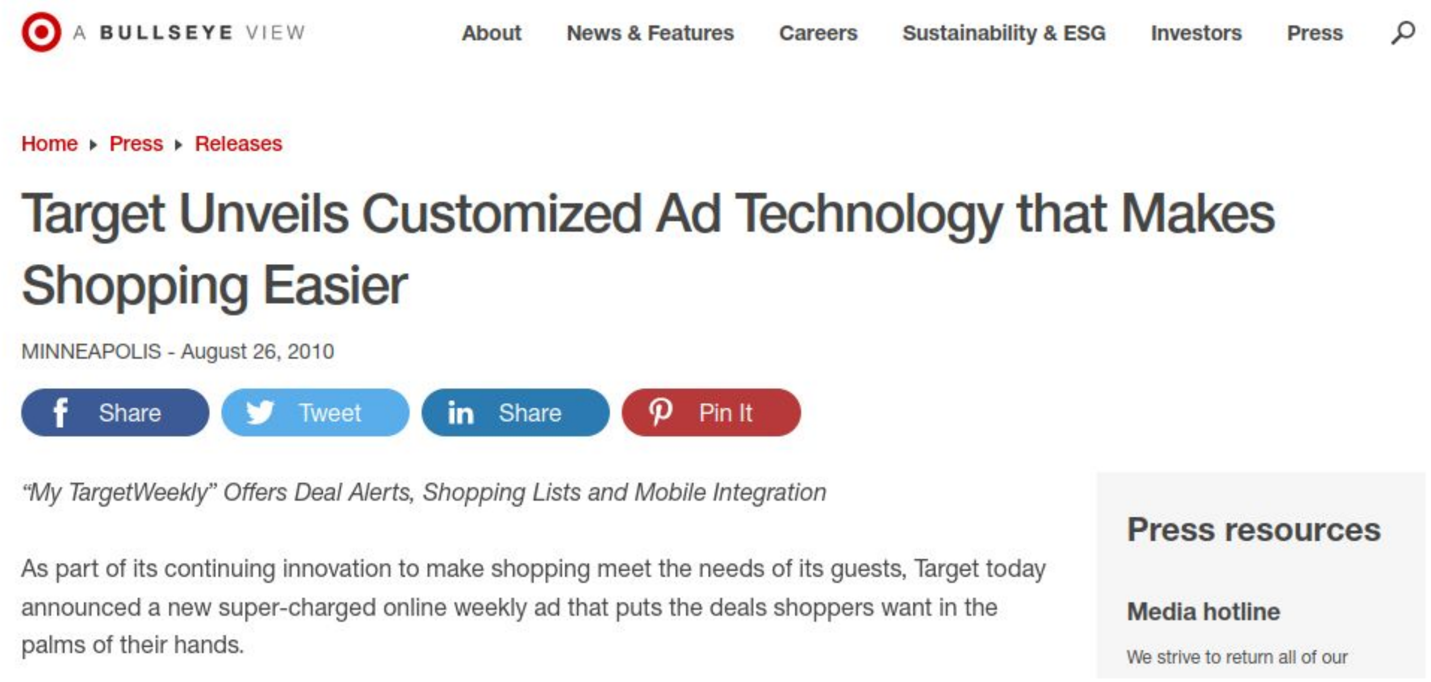
Arvind Narayanan and Vitaly Shmatikov

The University of Texas at Austin

February 5, 2008

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Still…



Home ▸ Press ▸ Releases

Target Unveils Customized Ad Technology that Makes Shopping Easier

MINNEAPOLIS - August 26, 2010

"My TargetWeekly" Offers Deal Alerts, Shopping Lists and Mobile Integration

As part of its continuing innovation to make shopping meet the needs of its guests, Target today announced a new super-charged online weekly ad that puts the deals shoppers want in the palms of their hands.

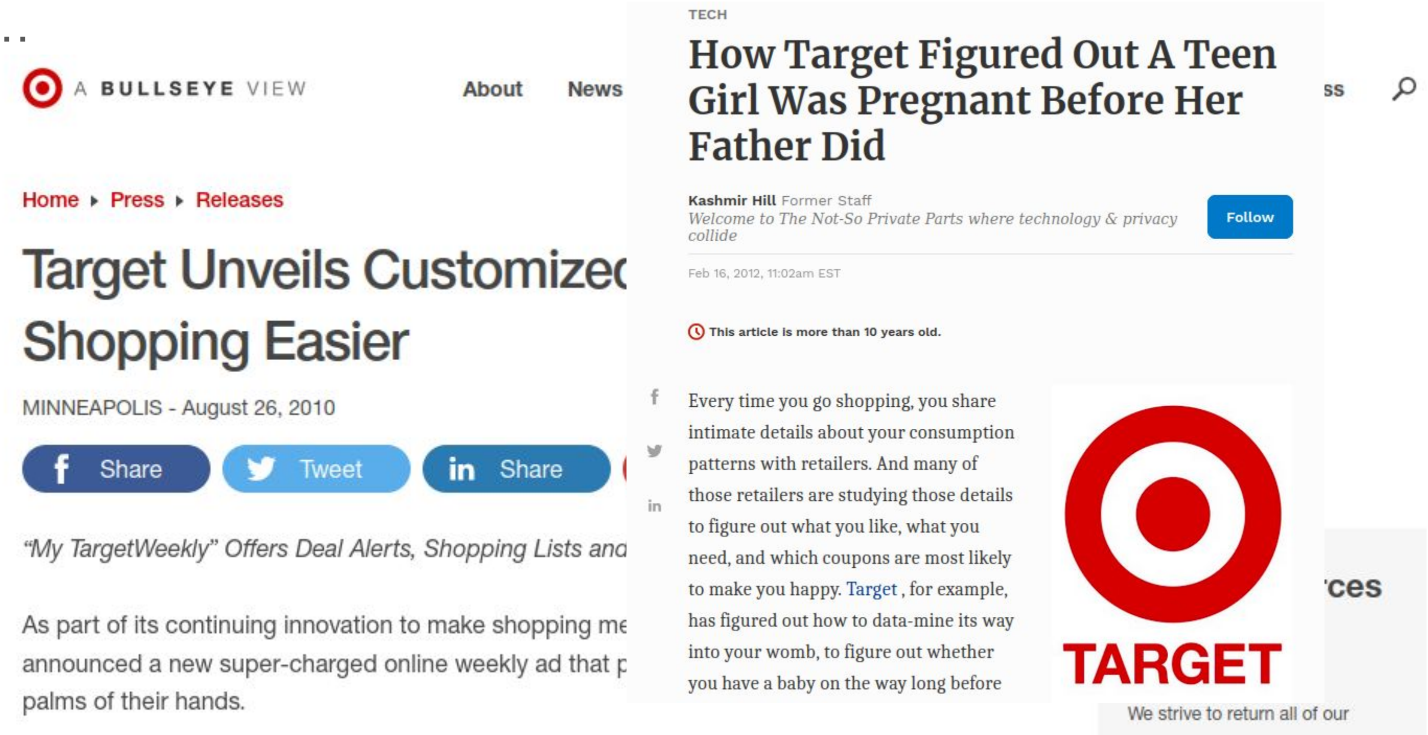
Press resources

Media hotline

We strive to return all of our

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Still…



Target Unveils Customizec
Shopping Easier

MINNEAPOLIS - August 26, 2010

“My TargetWeekly” Offers Deal Alerts, Shopping Lists anc

As part of its continuing innovation to make shopping me
announced a new super-charged online weekly ad that p
palms of their hands.

'ces

We strive to return all of our

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

Kashmir Hill Former Staff
Welcome to The Not-So Private Parts where technology & privacy collide

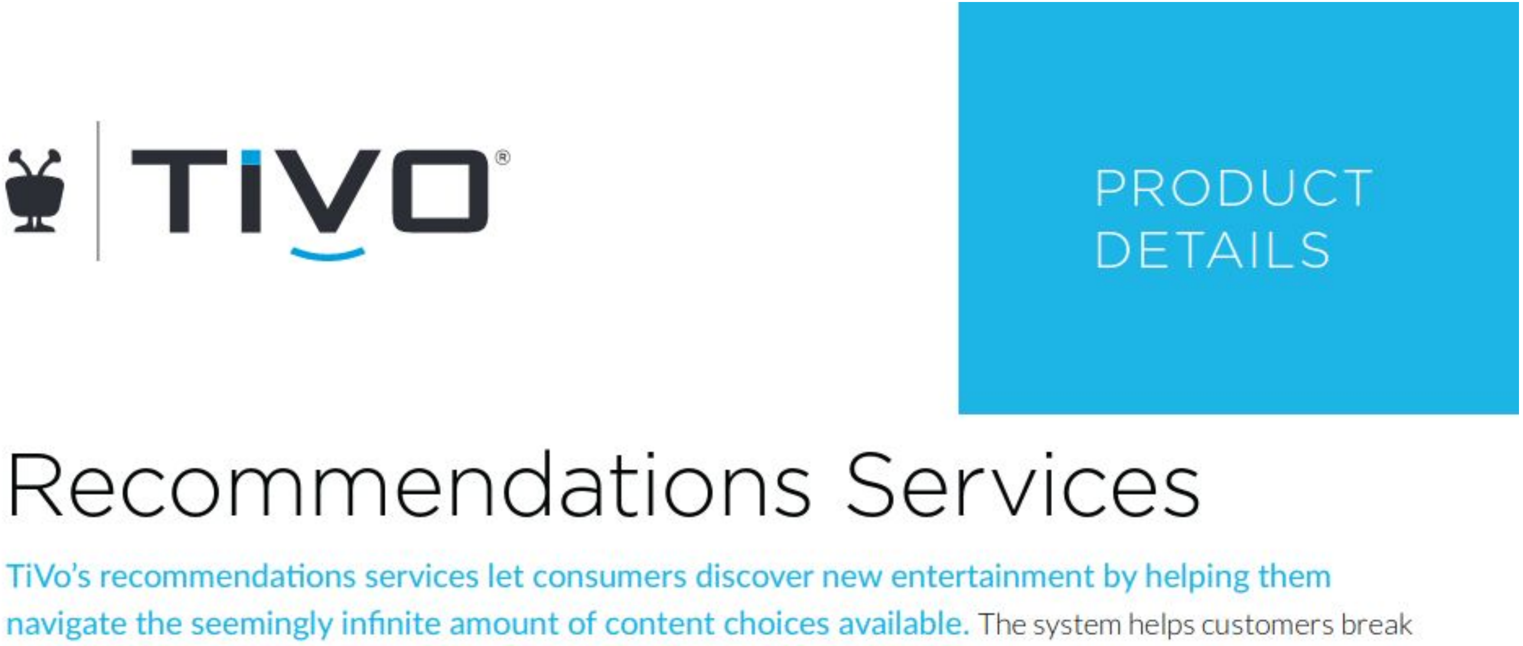
Feb 16, 2012, 11:02am EST

This article is more than 10 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 , 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

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Still…



TiVo®

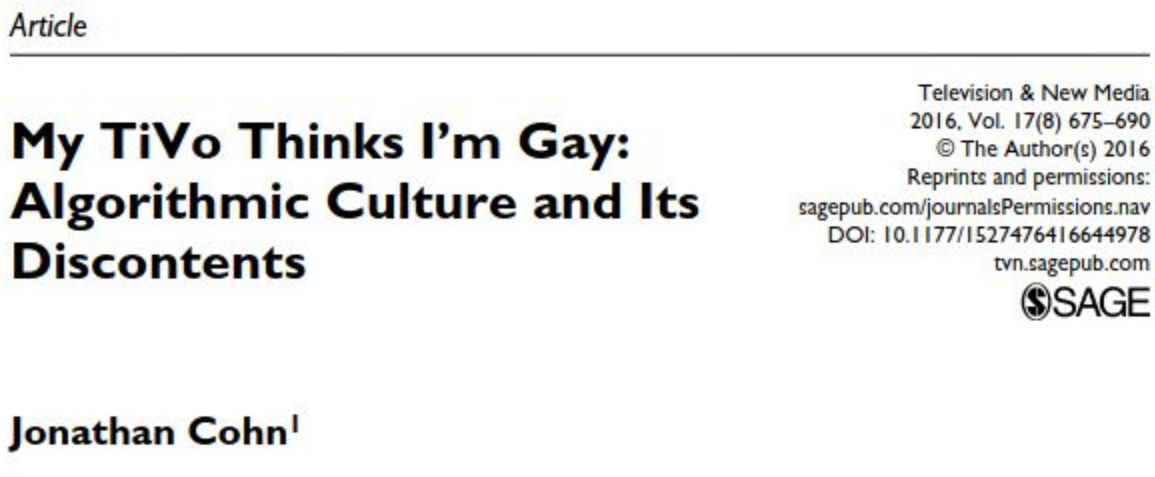
PRODUCT DETAILS

Recommendations Services

TiVo’s recommendations services let consumers discover new entertainment by helping them navigate the seemingly infinite amount of content choices available. The system helps customers break

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Still…

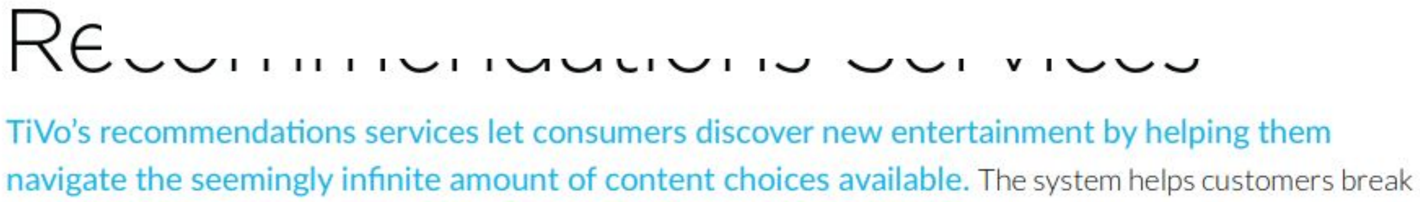


Article

My TiVo Thinks I'm Gay: Algorithmic Culture and Its Discontents

Television & New Media
2016, Vol. 17(8) 675–690
© The Author(s) 2016
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1527476416644978
tvn.sagepub.com
SAGE

Jonathan Cohn[1]



Recommendations Services

TiVo's recommendations services let consumers discover new entertainment by helping them navigate the seemingly infinite amount of content choices available. The system helps customers break

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Should the algorithm see our data?

PEEPING THROUGH THE HOLE: RECOMMENDER SYSTEMS

Should the algorithm see our data? The lesser, the better!

In general, **the more information we have, the more we have to protect** (+ keeping the data of the users requires us to “waste” resources)

PEEPING THROUGH THE WINDOW: DATABASES

Privacy risks regarding databases arise* when we **release data**:

by limiting which data we release (an in what form) we can somehow minimize the privacy risks…

*malevolent employees or other agents might pose a risk as well, but this risks do not directly concern privacy as we intend it

PEEPING THROUGH THE WINDOW: DATABASES

Microdata protection:

- mask the data
- perturbate the data
- generate new synthetic data

PEEPING THROUGH THE WINDOW: DATABASES

Privacy analysis:

- k-anonymity: ensure that, for each possible combination of attributes' values, k users have it.

- l-diversity, t-closeness: not only k diverse tuples, but also with diverse relevant attributes

PEEPING THROUGH THE WINDOW: DATABASES

Differential privacy:

- perturbate the data so that observations for a single user are “useless”, but aggregated statistics remain valid.

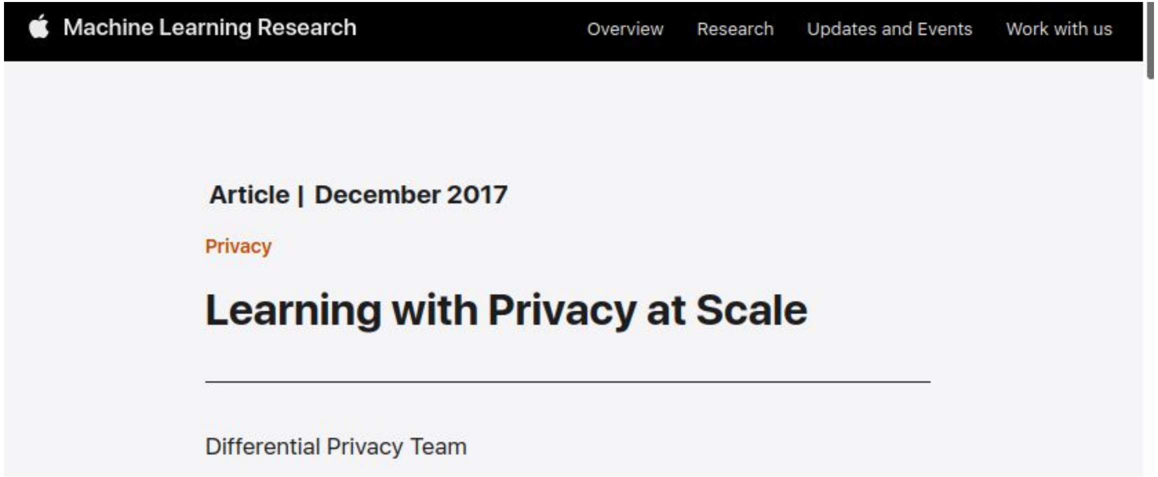
Very popular!

PEEPING THROUGH THE WINDOW: DATABASES

Differential privacy:

- perturbate the data so that observations for a single user are “useless”, but aggregated statistics remain valid.

Very popular!



Article | December 2017

Privacy

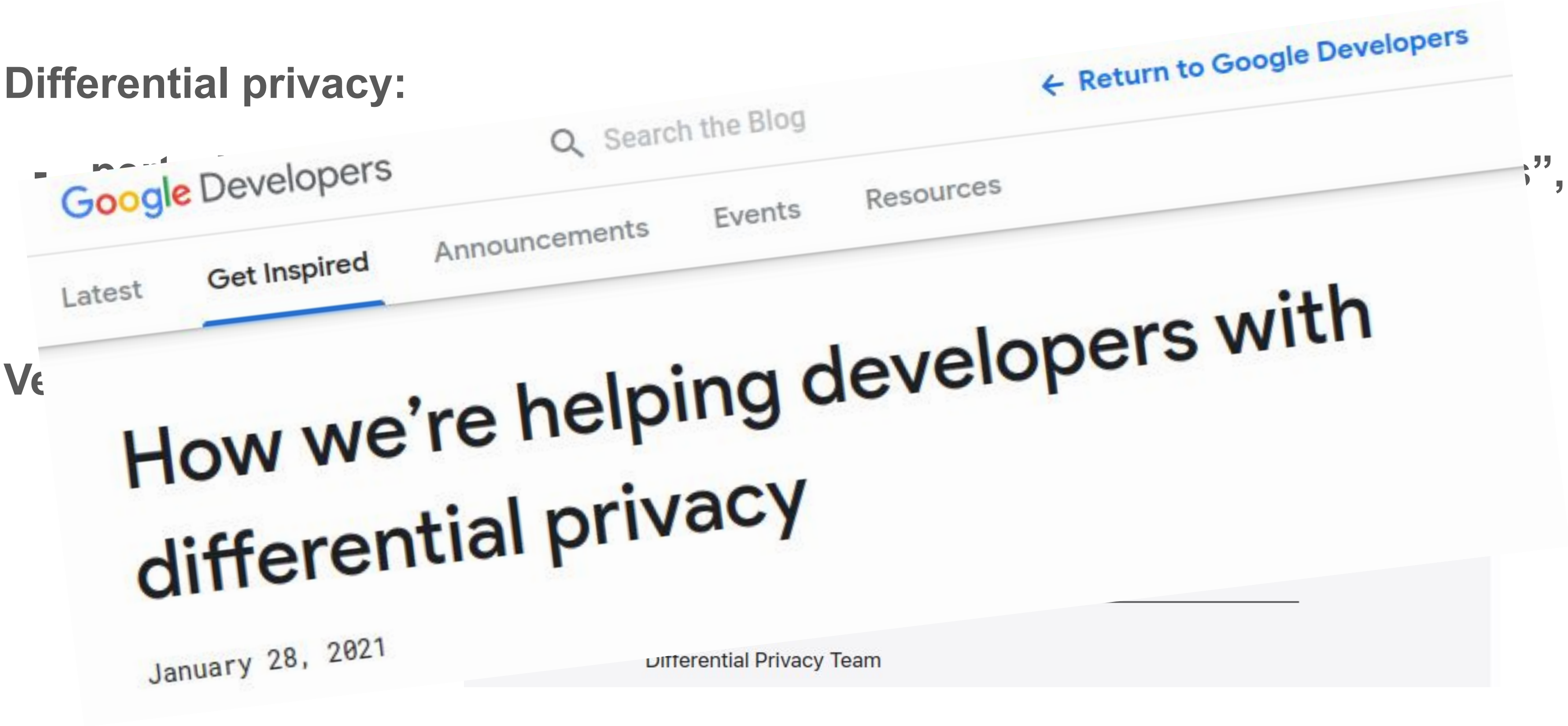
Learning with Privacy at Scale

Differential Privacy Team

PEEPING THROUGH THE WINDOW: DATABASES

Differential privacy:

Ve



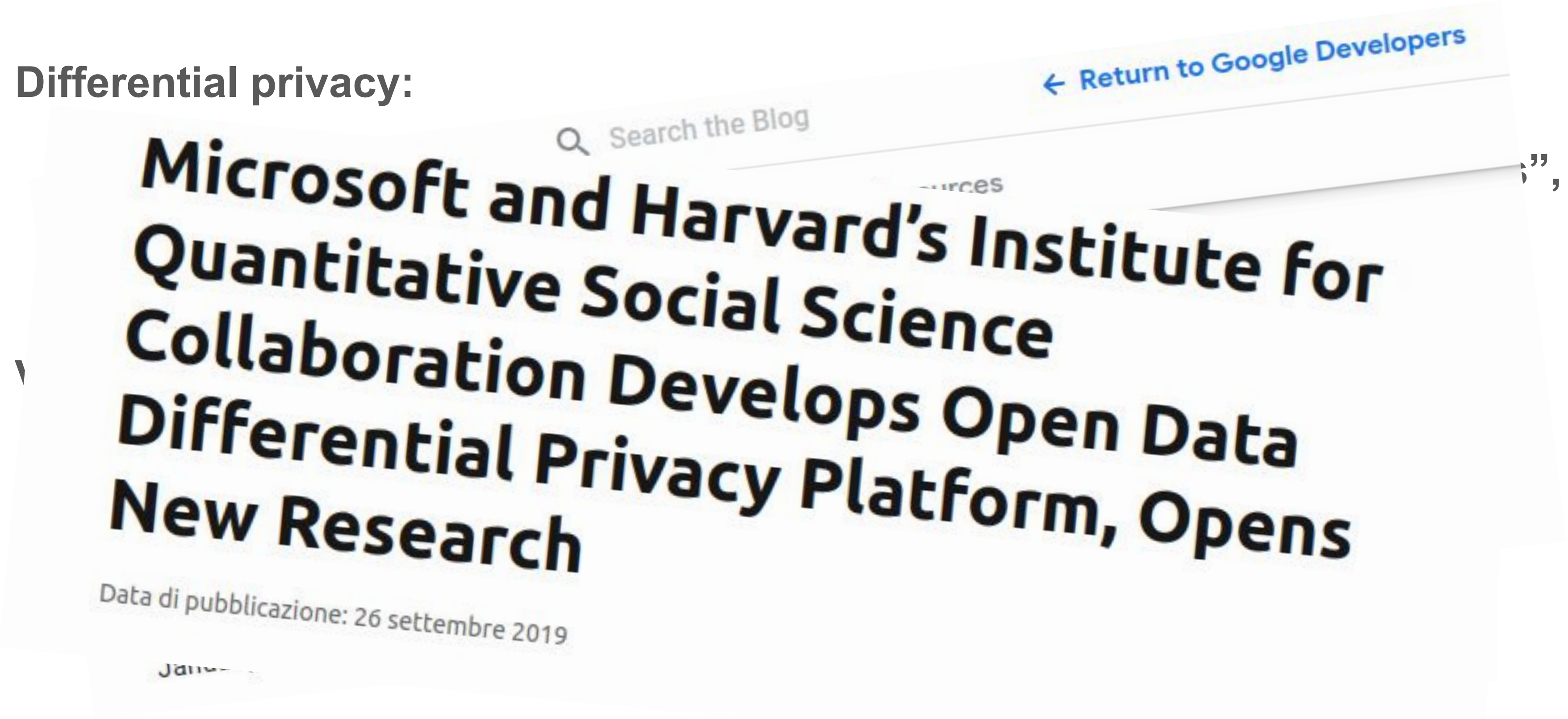
How we’re helping developers with differential privacy

January 28, 2021

Differential Privacy Team

PEEPING THROUGH THE WINDOW: DATABASES

Differential privacy:



Microsoft and Harvard’s Institute for Quantitative Social Science Collaboration Develops Open Data Differential Privacy Platform, Opens New Research

Data di pubblicazione: 26 settembre 2019