Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning

Hamza Harkous (EPFL)

Kassem Fawaz (U. Wisconsin)

Rémi Lebret (EPFL)

Florian Schaub (U. Michigan)

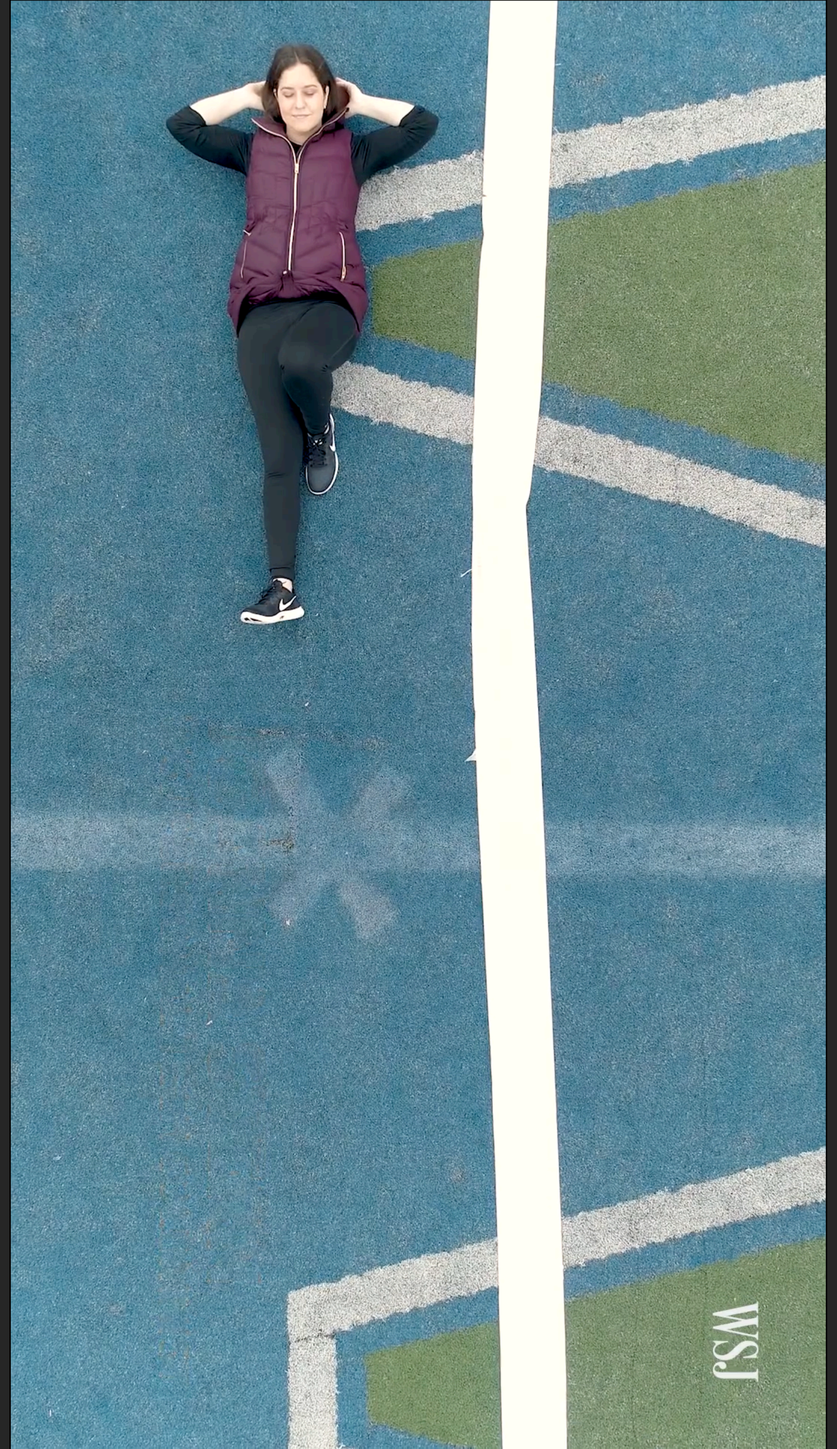
Kang G. Shin (U. Michigan)

Karl Aberer (EPFL)

pribot.org

hamzaharkous.com

Problem?



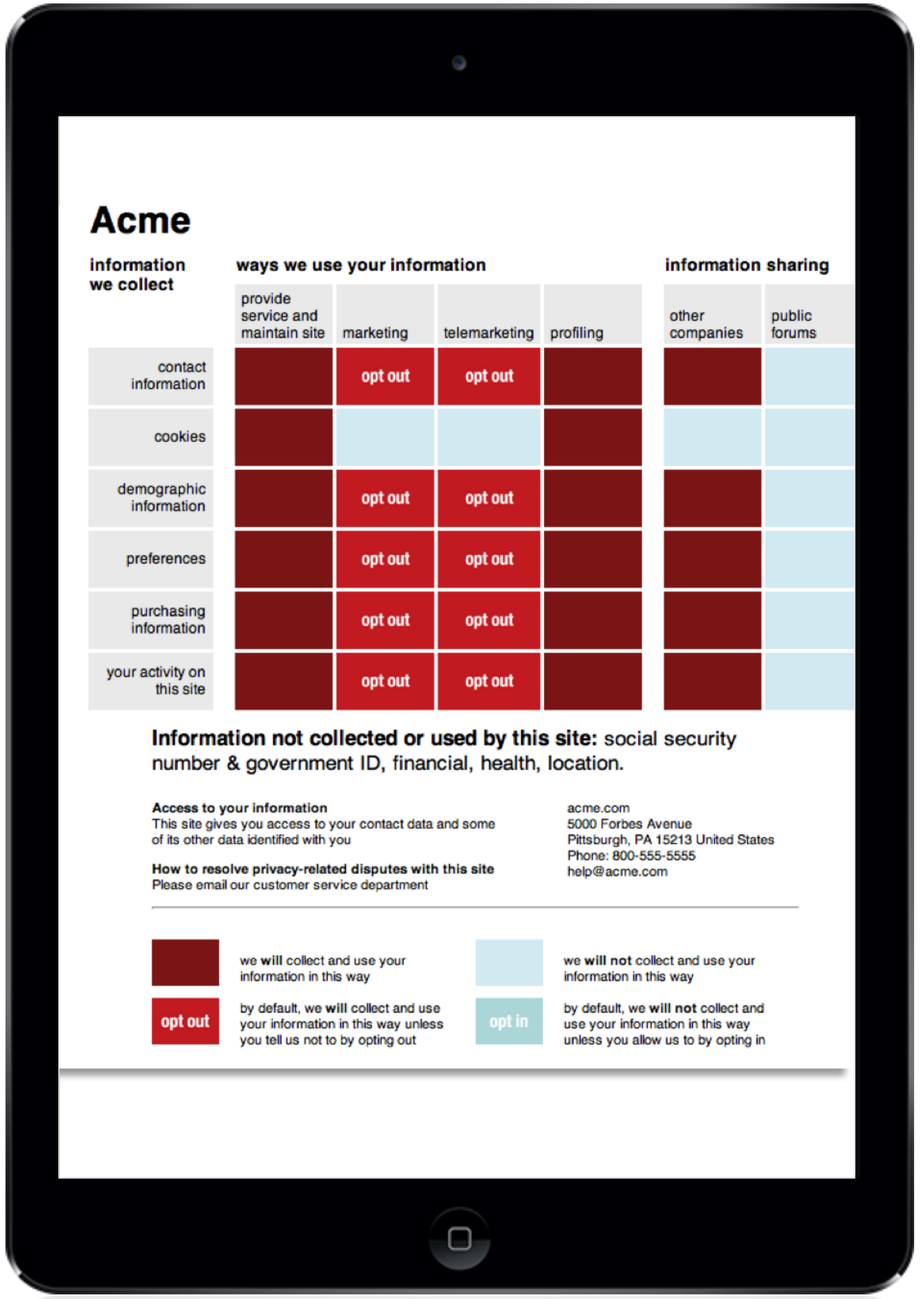
PRIVACY POLICIES ARE LONG AND COMPLEX

201 hours per year on average to read policies of services we encounter*

*(McDonald and Cranor -2008)

APPROACHES SO FAR?

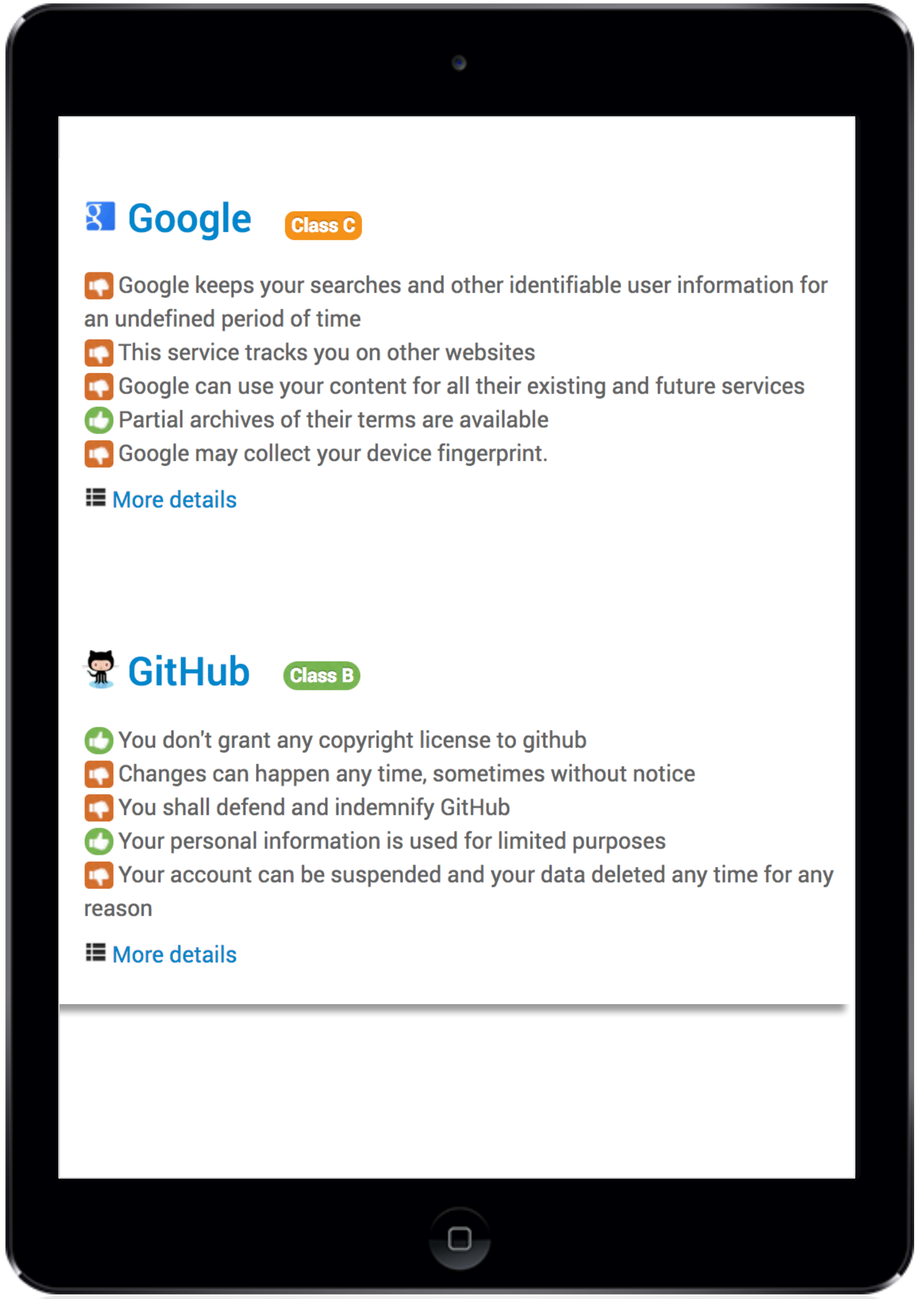
Put more lawyers on the task.



Standardization

- A Nutrition Label for privacy
- Required providers to act
- **Surprise**: They didn't.

Kelley et al., "A nutrition label for privacy." SOUPS'09



Crowdsourcing

- TOSDR.org
- Limited by volunteers' availability
- Available for ~100 policies
- Unstructured → can only be used for limited automated labeling*

*Zimmeck and Bellovin , "Privee: An Architecture for Automatically Analyzing Web Privacy Policies". USENIX Security 2014

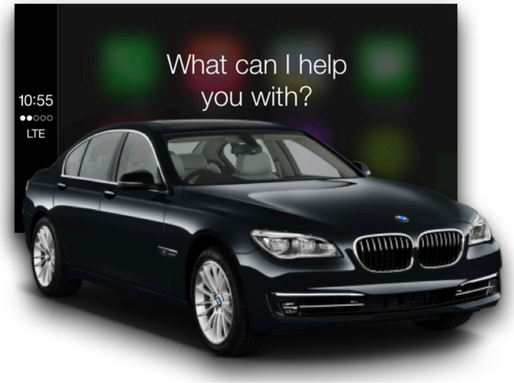
Manual work doesn’t scale.

• • •

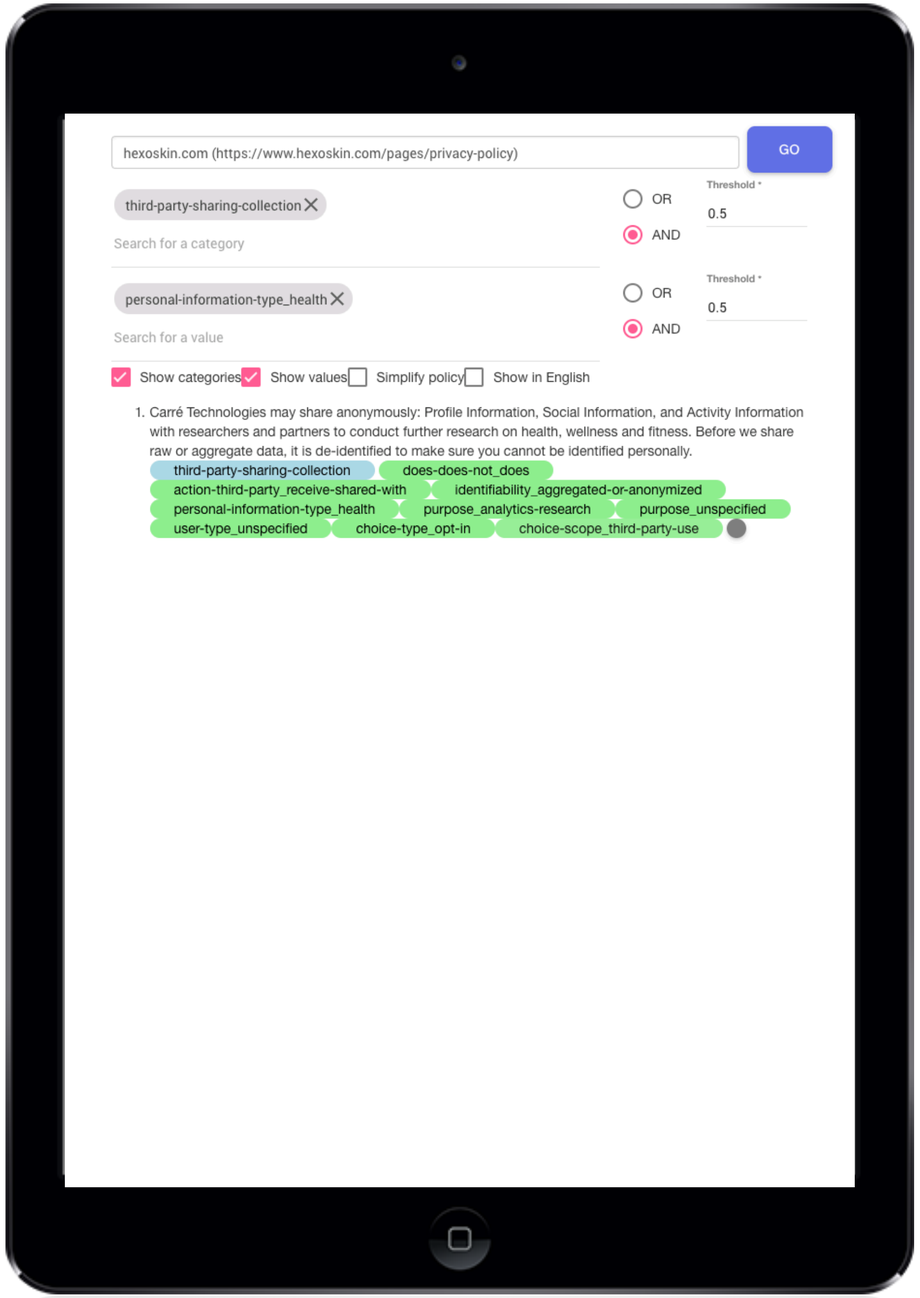
Fails to cope with emerging technologies.

Voice-Activated Devices

- Read the whole policy?
- Show tables on small screens?



Unstructured Query (User Questions)



Regulation Compliance (e.g. GDPR)

Find Statements About Health Data Sharing

Get Segments such that

Category: third party sharing

personal information type: health information

Structured Query

Solution

POLISIS

Unified Framework for Privacy Policies Analysis

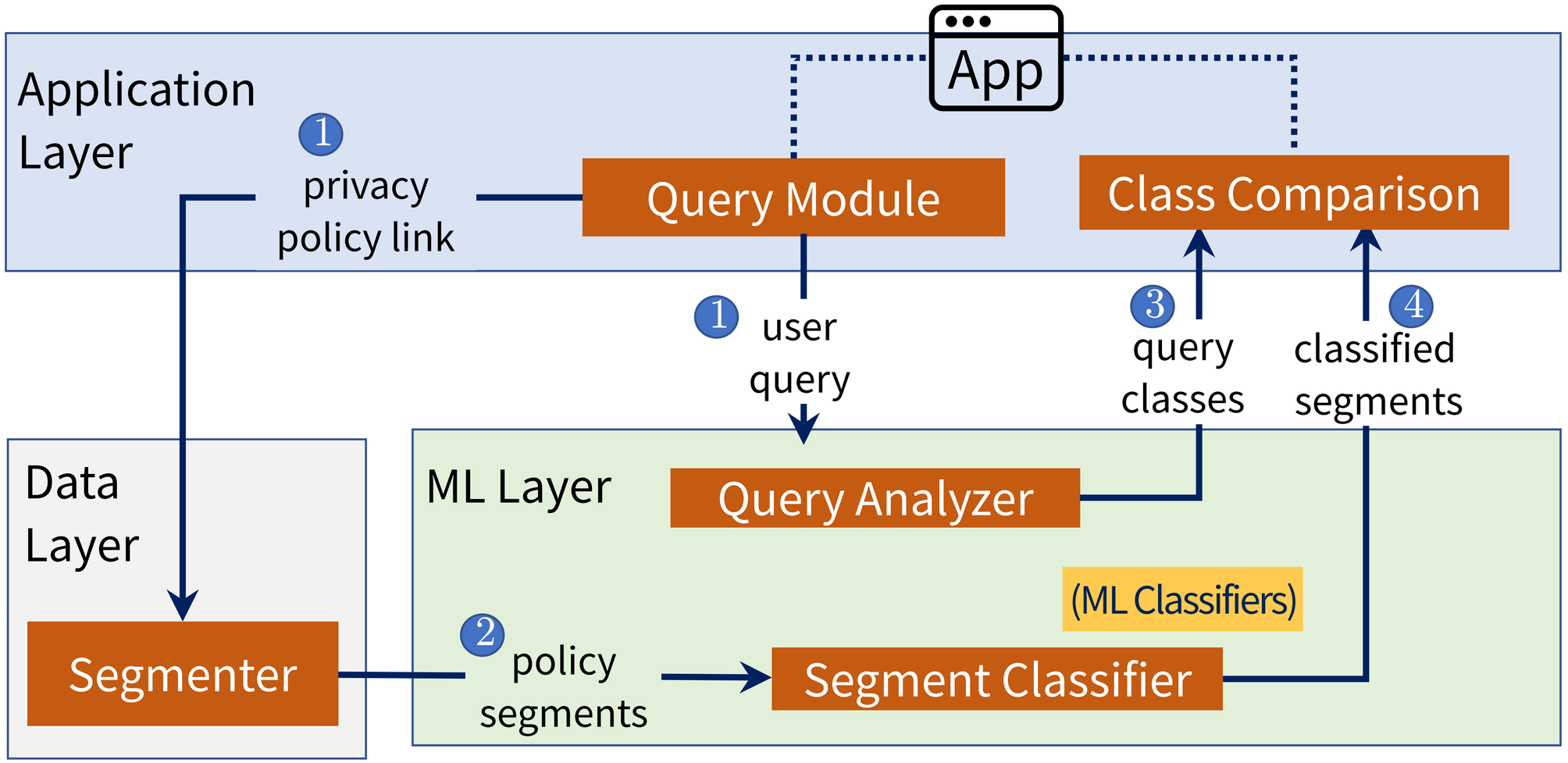
Once we automate policies' analysis, we can create a new interface for millions of policies with a single program.

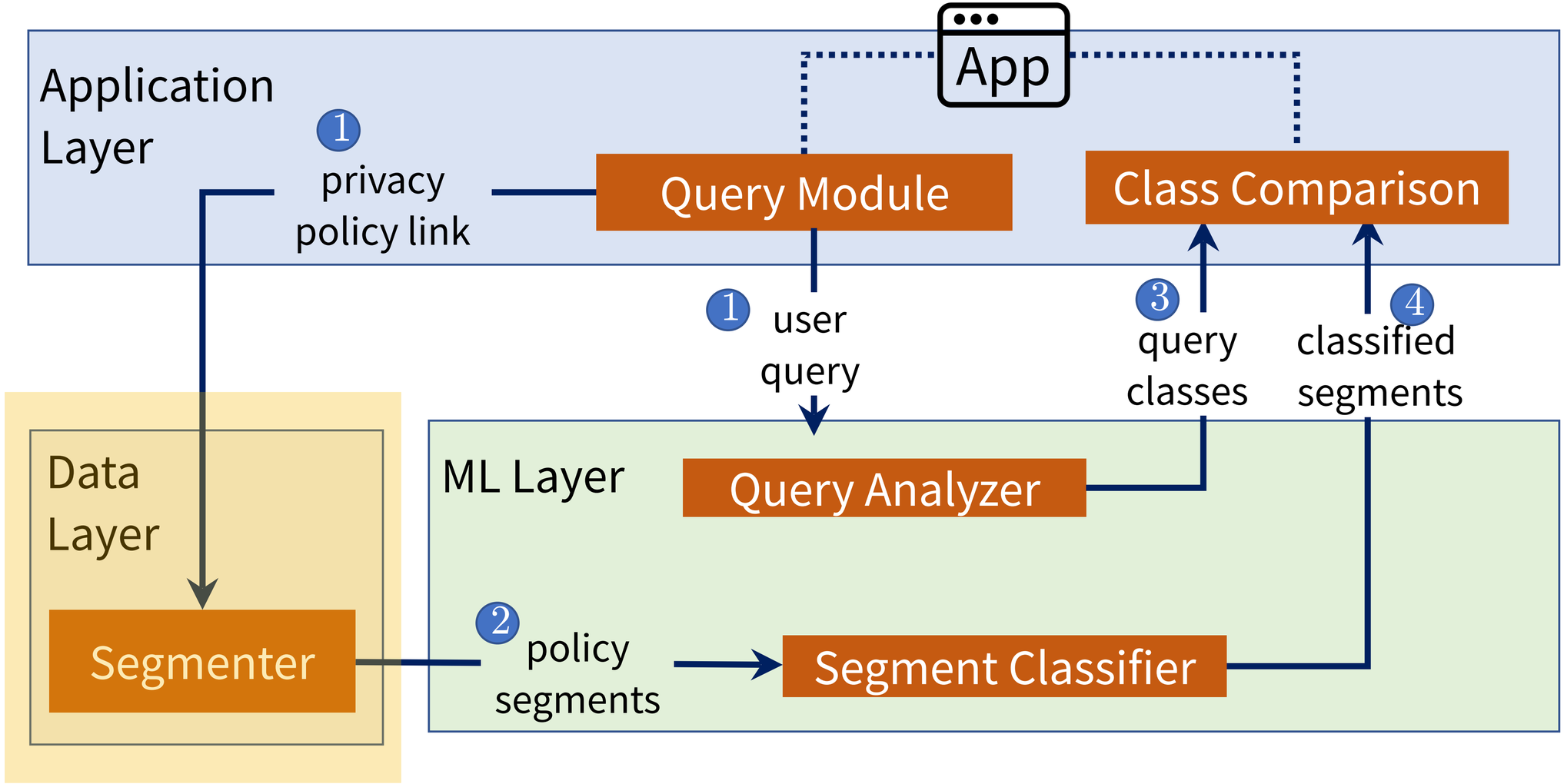
Structured Queries

Unstructured Queries

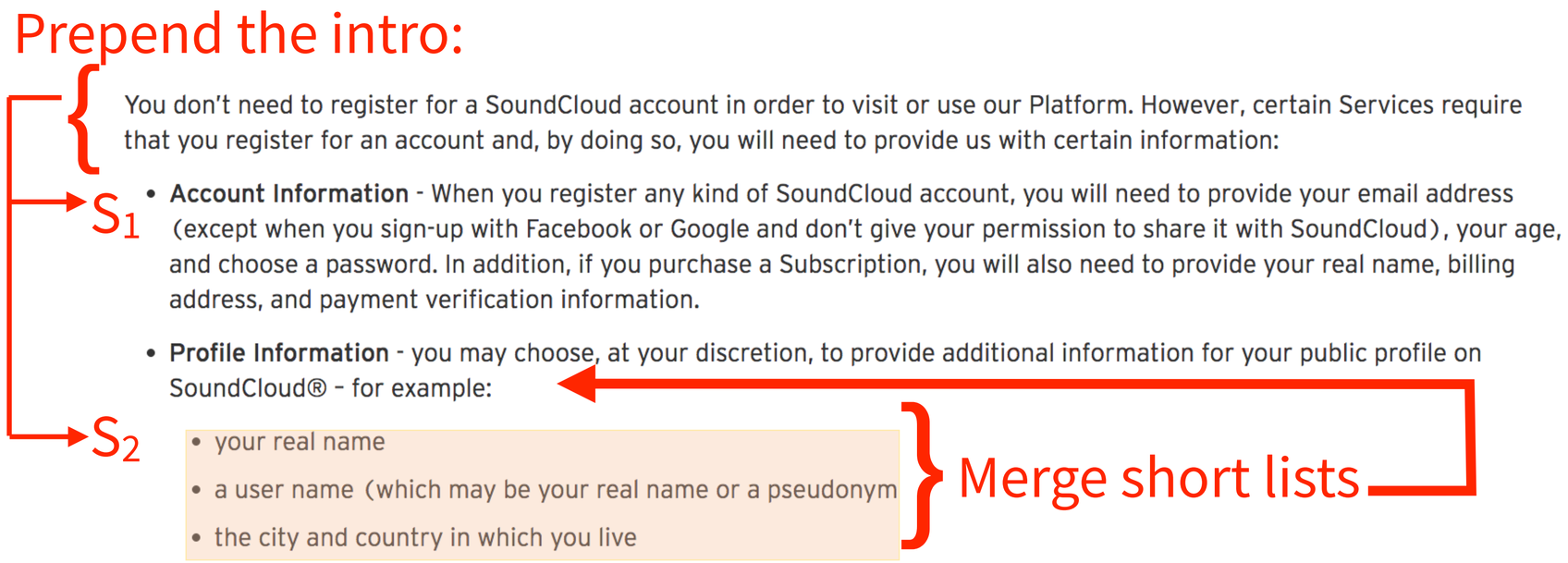
pribot.org

Framework


App
Application Layer
1
privacy policy link
Query Module
Class Comparison
1
user query
3
query classes
4
classified segments
Data Layer
ML Layer
Query Analyzer
(ML Classifiers)
Segmenter
2
policy segments
Segment Classifier

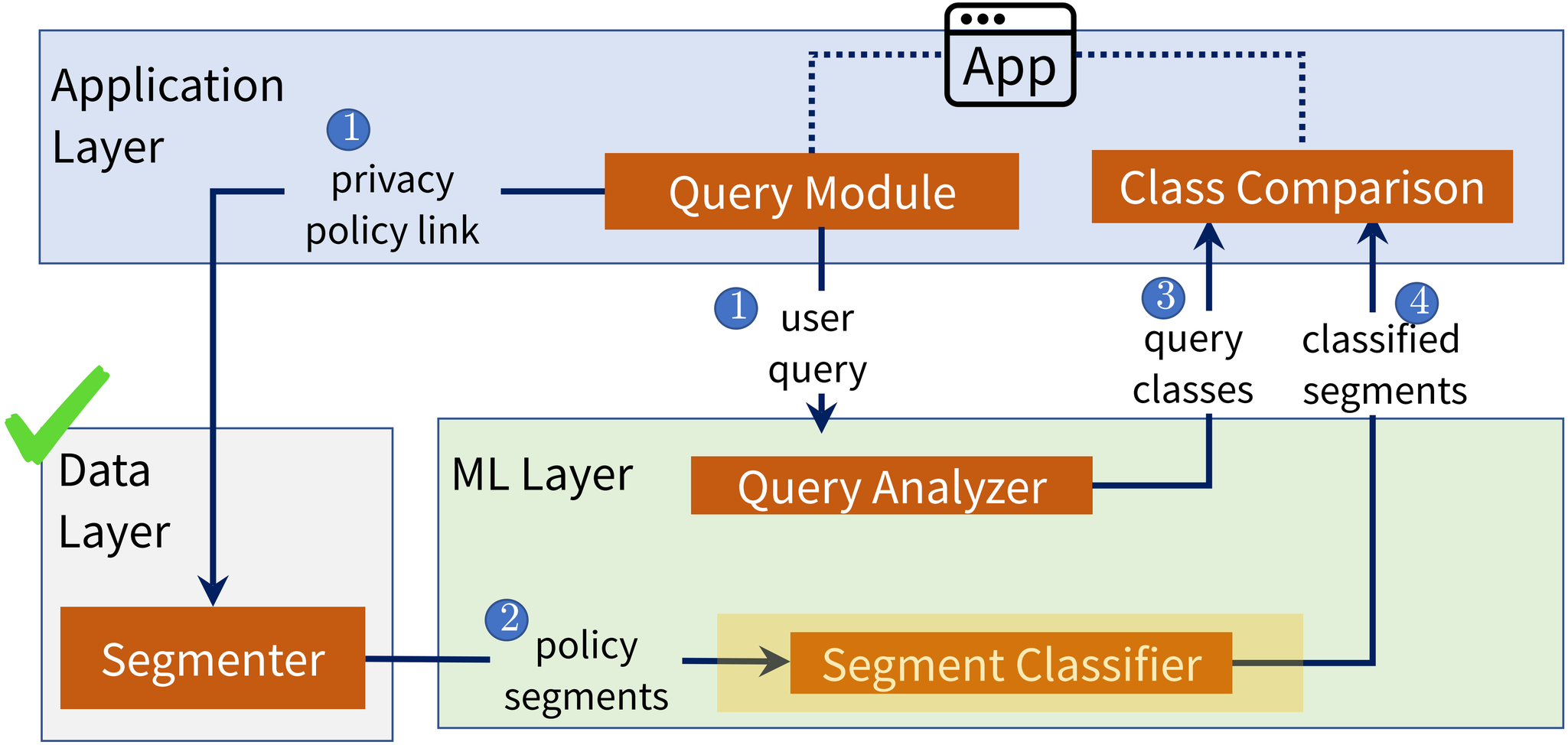

Application Layer
App
1
privacy policy link
Query Module
Class Comparison
1
user query
3
query classes
4
classified segments
Data Layer
ML Layer
Query Analyzer
Segmenter
2
policy segments
Segment Classifier

Policy Segmenter



HTML Preprocessing → Coarse Segmentation according to HTML Tags → Text Segmentation into semantically coherent segments*

*Glavas et al., "Unsupervised Text Segmentation Using Semantic Relatedness Graphs", ACL 2016


Application Layer
App
1
privacy policy link
Query Module
Class Comparison
1
user query
3
query classes
4
classified segments
Data Layer
ML Layer
Query Analyzer
Segmenter
2
policy segments
Segment Classifier

EXAMPLE

Intel's Privacy Policy

We may need to retain certain information for recordkeeping purposes, as required under applicable legal obligations, and/or to complete any transactions that you began prior to requesting such change or deletion (…) Some of your information may remain within our systems and other records, in compliance with applicable law.

EXPERT ANNOTATIONS

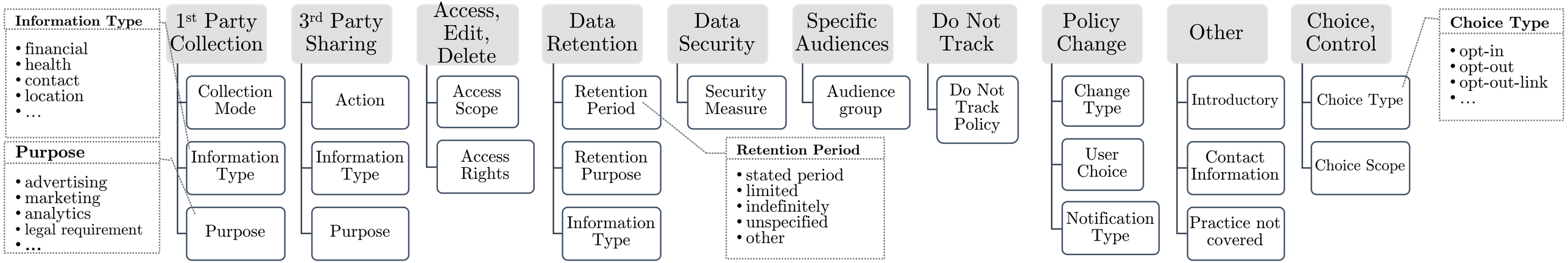
Data Retention

- purpose
  - legal requirement
- info type
  - generic

*OPP-115 dataset by Wilson et al., ACL 2016

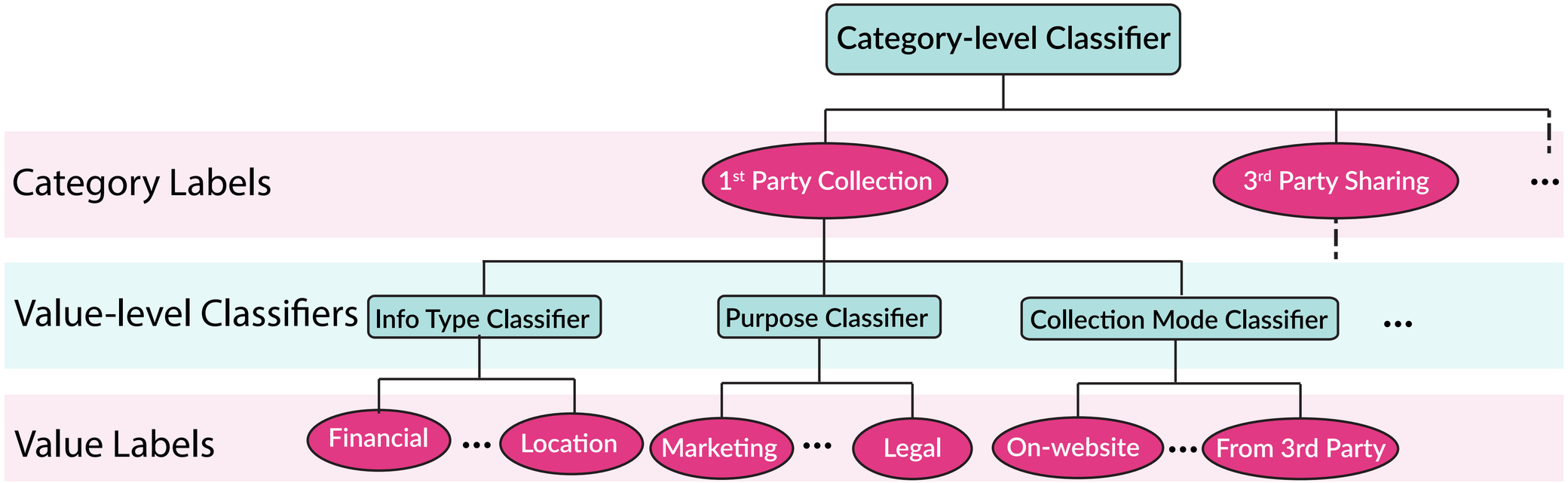
Online Privacy Policies Dataset

- 115 annotated policies
- 23K annotations



Hierarchical Data  Hierarchical Architecture

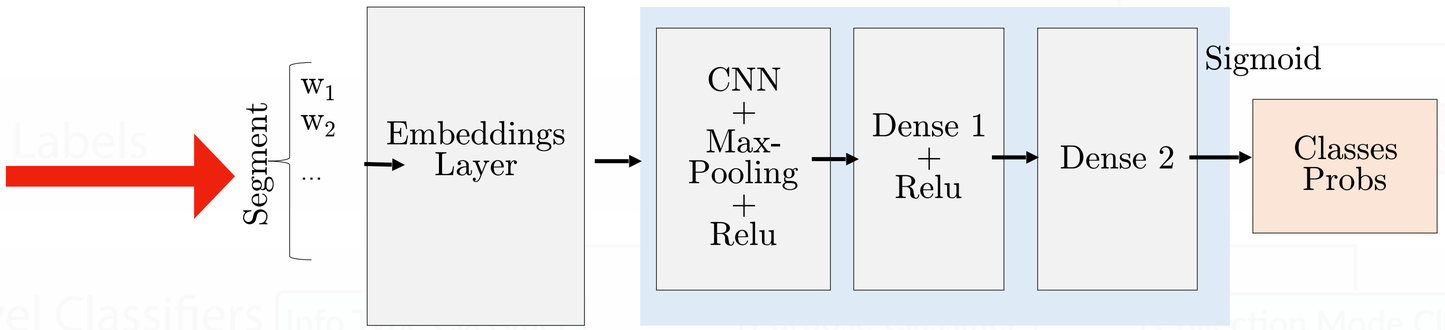
Hierarchical Architecture



Hierarchical Architecture

130,000 privacy policies from Play Store to train our custom word embeddings

Similar architecture for the 21 classifiers

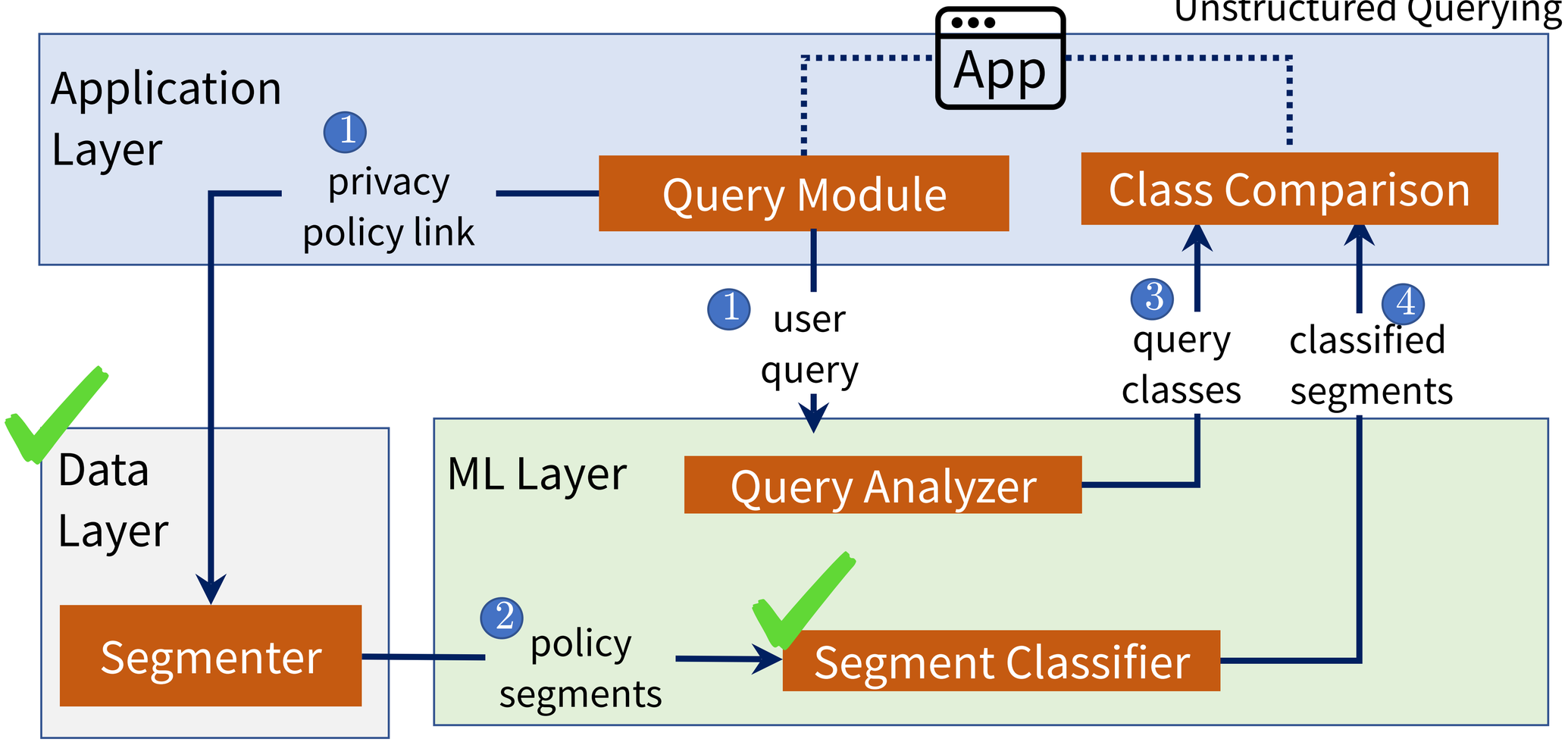


Loss function: multi-label cross entropy

Embeddings size: 300, Number of filters: 200, Filter Size: 3, Dense Layer Size: 100, Batch Size: 40

Structured Querying

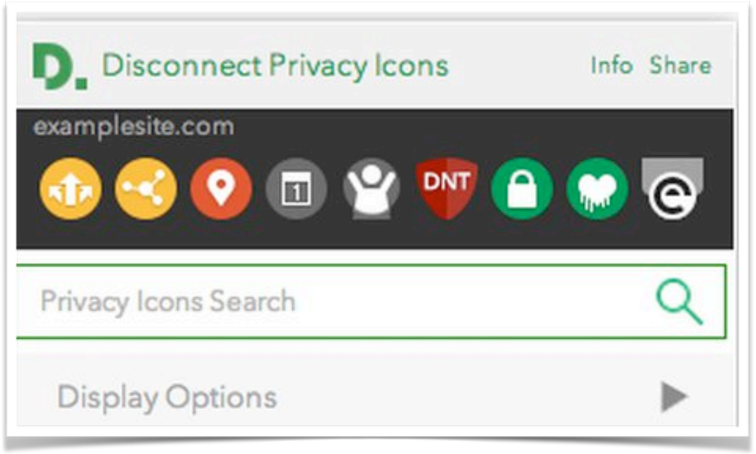
Unstructured Querying



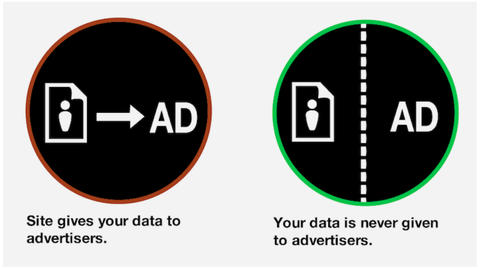
Structured Querying

Privacy Icon Assignment as a Case Study

Automated Privacy Icons Assignment



[1]



[2]

TRUSTe & Disconnect Introduce Visual Icons to Help Consumers Understand Privacy Policies

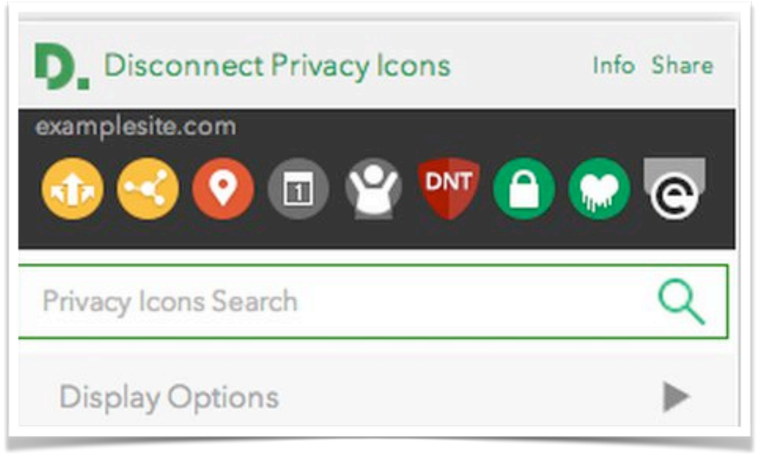
June 23, 2014

Today, privacy innovators TRUSTe & Disconnect have launched Privacy Icons software to help consumers easily understand website privacy policies and how websites are handling their data.

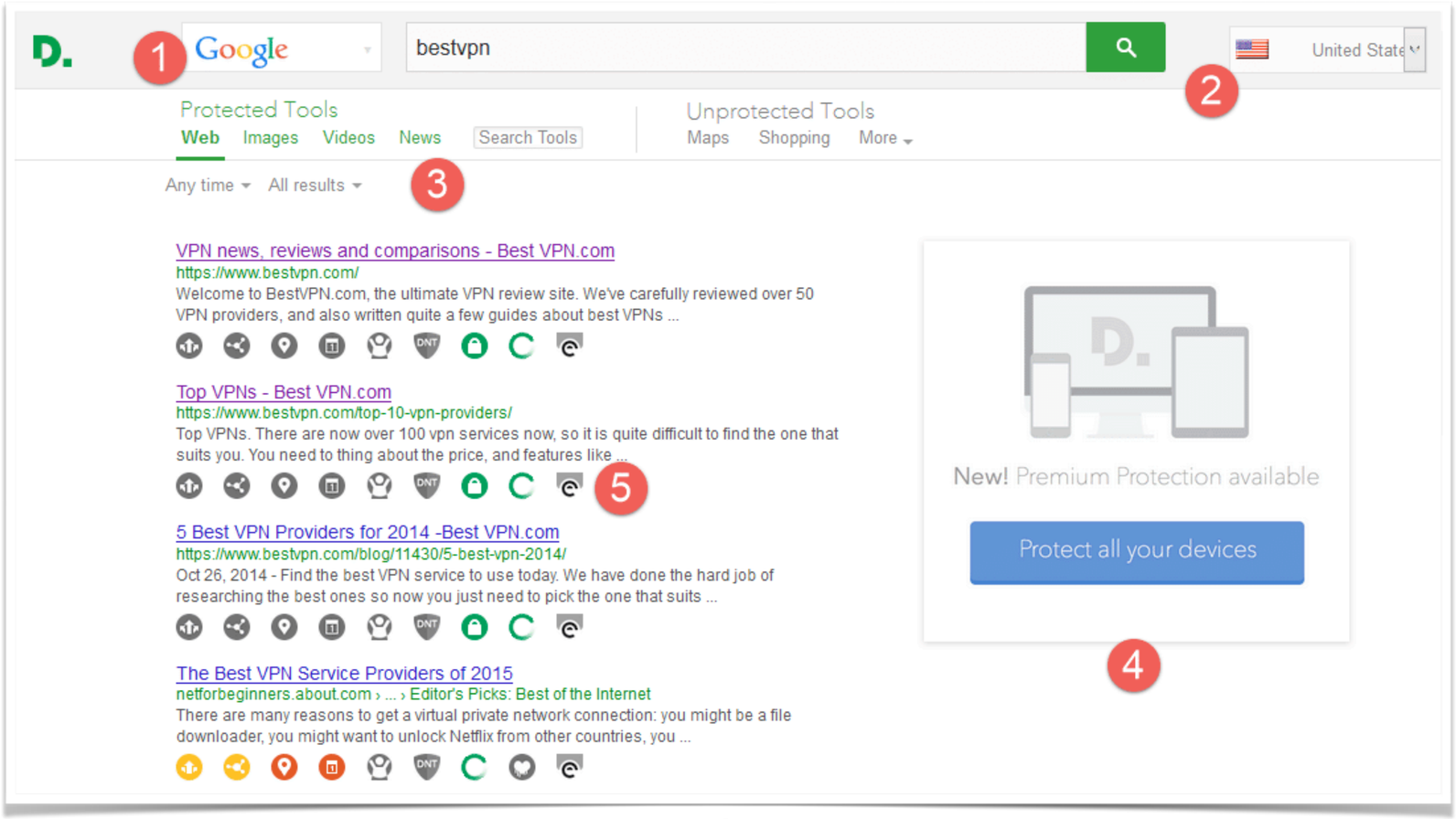
Consumers want to know how websites are using their privacy and data, but they often do not have the time or patience to read existing privacy policies, which are typically quite long and complex. According to The TRUSTe Privacy Index, the average privacy policy is 2,464 words long and takes about 10 minutes to read.

[1] https://web.archive.org/web/20170709022651/https://disconnect.me/icons
[2] https://wiki.mozilla.org/Privacy_Icons

Deployment of Disconnect Icons



Chrome Extension



Web App

Discontiuned in 2017 ☹

Disconnect Icons Description

Expected Use

Does this website's privacy policy disclose whether data it collects about you is used in ways other than you would reasonably expect given the site's service?

Red = Yes, without choice to opt-out. Or, undisclosed.

Yellow = Yes, with choice to opt-out.

Green = No.

Gray = Info unavailable.

Expected Collection

Does this website's privacy policy disclose whether it allows other companies like ad providers and analytics firms to track users on the site?

Red = Yes, without choice to opt-out. Or, undisclosed.

Yellow = Yes, with choice to opt-out.

Green = No.

Gray = Info unavailable.

Precise Location

Does this website's privacy policy disclose whether the site or service tracks a user's actual geolocation?

Red = Yes, possibly without choice.

Yellow = Yes, with choice.

Green = No.

Gray = Info unavailable.

Data Retention

Does this website's privacy policy disclose how long they retain your personal data?

Red = No data retention policy.

Yellow = 12+ months.

Green = 0-12 months.

Gray = Info unavailable.

Do Not Track

Does this website comply with a user's Do Not Track browser preference?

Green = Yes.

Gray = Info unavailable.

Children Privacy

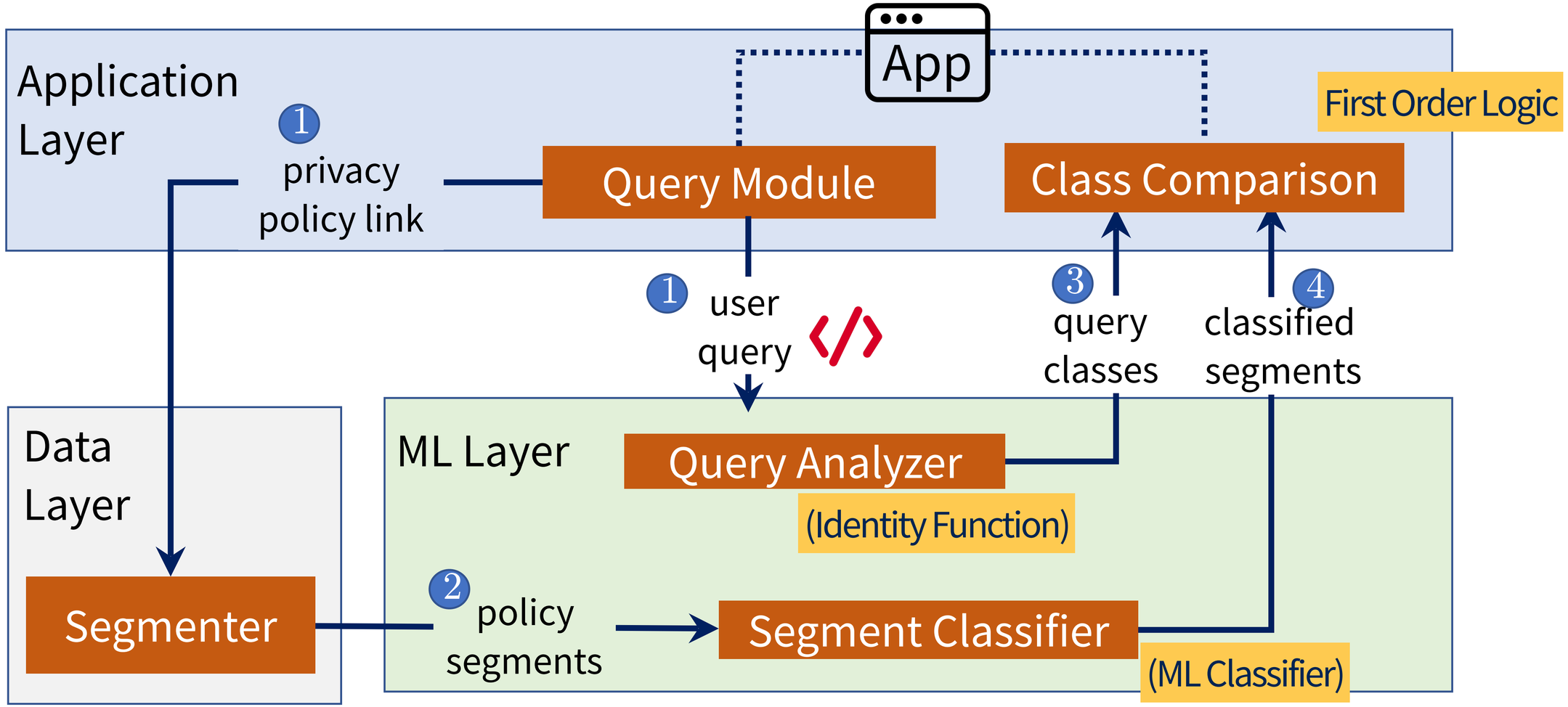
Has this website received TRUSTe's Children's Privacy Certification?

Green = Yes.

Gray = No.

Structured Query

```
Get Segments such that
Category: third party sharing
purpose:  advertising
```



Expected Collection

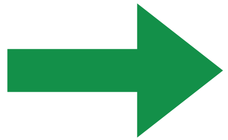
Does this website's privacy policy disclose whether it allows other companies like ad providers and analytics firms to track users on the site?

Red = Yes, without choice to opt-out. Or, undisclosed.

Yellow = Yes, with choice to opt-out.

Green = No.

Gray = Info unavailable.



Structured Query

```
color: yellow
category_filters:
 -
   fun: include_some_in_list
   lst: ['third-party-sharing-collection']

value_filters:
 -
    fun: include_some_in_list
    lst: ['purpose_advertising','purpose_analytics-research']

 -
    fun: include_some_in_list
    lst: ['action-third-party_track-on-first-party-website-app',
          'action-third-party_collect-on-first-party-website-app']

 -
    fun: include_some_in_list
    lst: ['choice-type_opt-out-link',
          'choice-type_opt-out-via-contacting-company']


decider:
  -
    fun: not_empty
```

Structured Query

```
-
  color: yellow
  category_filters:
   -
     fun: include_some_in_list
     lst: ['third-party-sharing-collection']

  value_filters:
   -
      fun: include_some_in_list
      lst: ['purpose_advertising','purpose_analytics-research']

   -
      fun: include_some_in_list
      lst: ['action-third-party_track-on-first-party-website-app',
            'action-third-party_collect-on-first-party-website-app']

   -
      fun: include_some_in_list
      lst: ['choice-type_opt-out-link',
            'choice-type_opt-out-via-contacting-company']


  decider:
    -
      fun: not_empty
```

88.4% accuracy

Same rules, on 50 policies from OPP-115 dataset

Icon assignment based on **law students' labels**

vs.

Icon assignment based on **Polisis' labels**

Icon	Accuracy
Exp. Use	92%
Exp. Collection	88%
Precise Location	84%
Data Retention	80%
Children Privacy	98%

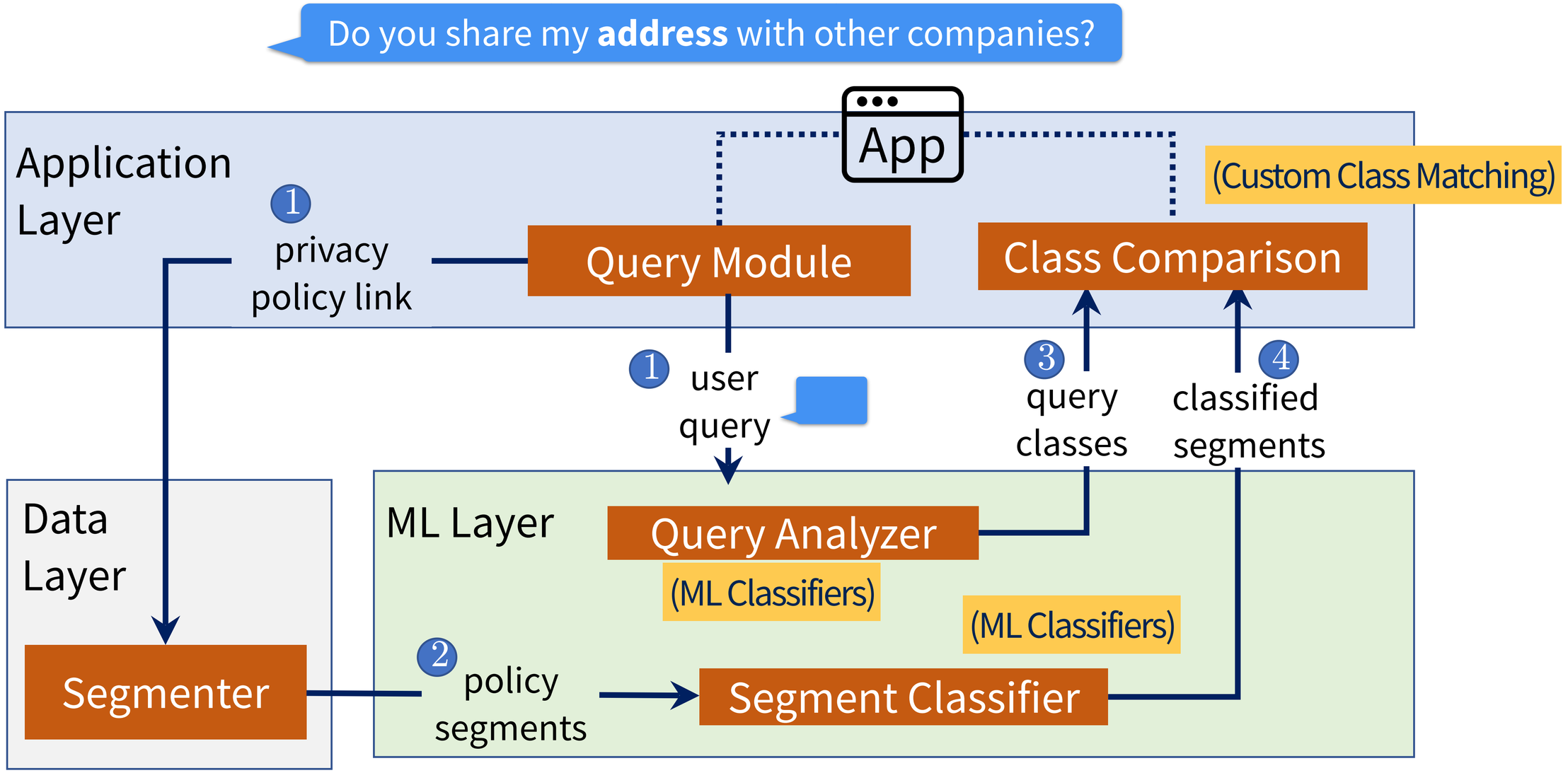
Unstructured Querying:
Answer Selection as a Case Study

Answer Selection

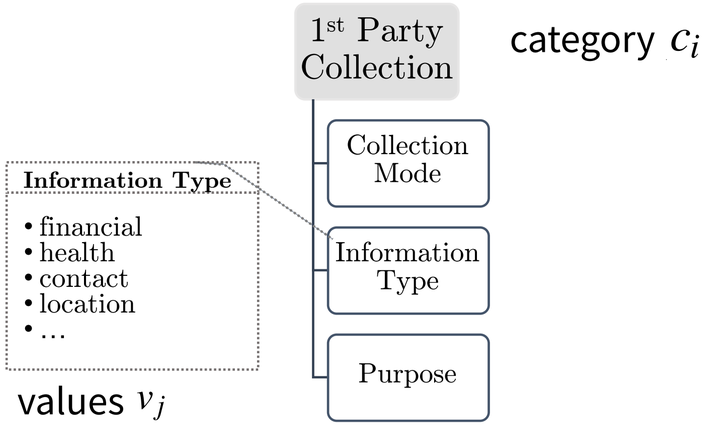
Do you share my **address** with other companies?

We will provide your **location** to third parties.

Unstructured Queries



Classifier/Analyzer



Answer vector

$$\boldsymbol{\alpha} = \{p(c_i|a)^2 \times p(v_j|a)\}$$

Query Vector

$$\boldsymbol{\beta} = \{p(c_i|q)^2 \times p(v_j|q)\}$$

3 classes for query q

4 classes for segment a

ML Layer

Query Analyzer

2 policy segments

Segment Classifier

Answer-Query Score

$$s(q,a) = \frac{\sum_i (\beta_i \times \min(\beta_i, \alpha_i))}{\sum_i \beta_i^2} \times cer(a)$$

prioritizes answers that include the question's classes with high probability (but not necessarily vice-versa)

Category Certainty Measure

$$cer(a) = 1 - (-\sum(p_n(c_i|a) \times \ln(p_n(c_i|a))/\ln(|\mathcal{C}|))$$

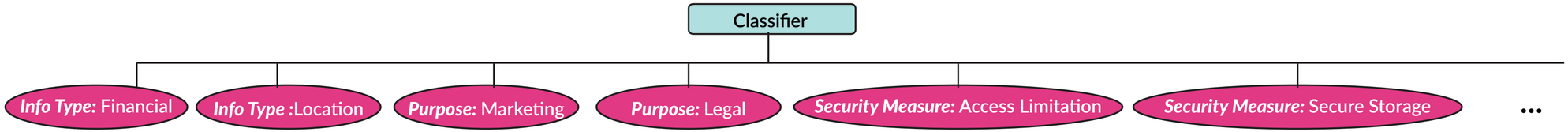
Evaluation Baselines

1- BM25 Model (Term-matching state of the art)

- Compute IDF values on the corpus of 130,000 policies

2- Semantic Vector Model

- Flat hierarchy across all values, CNN classifier



3- Random answer Model

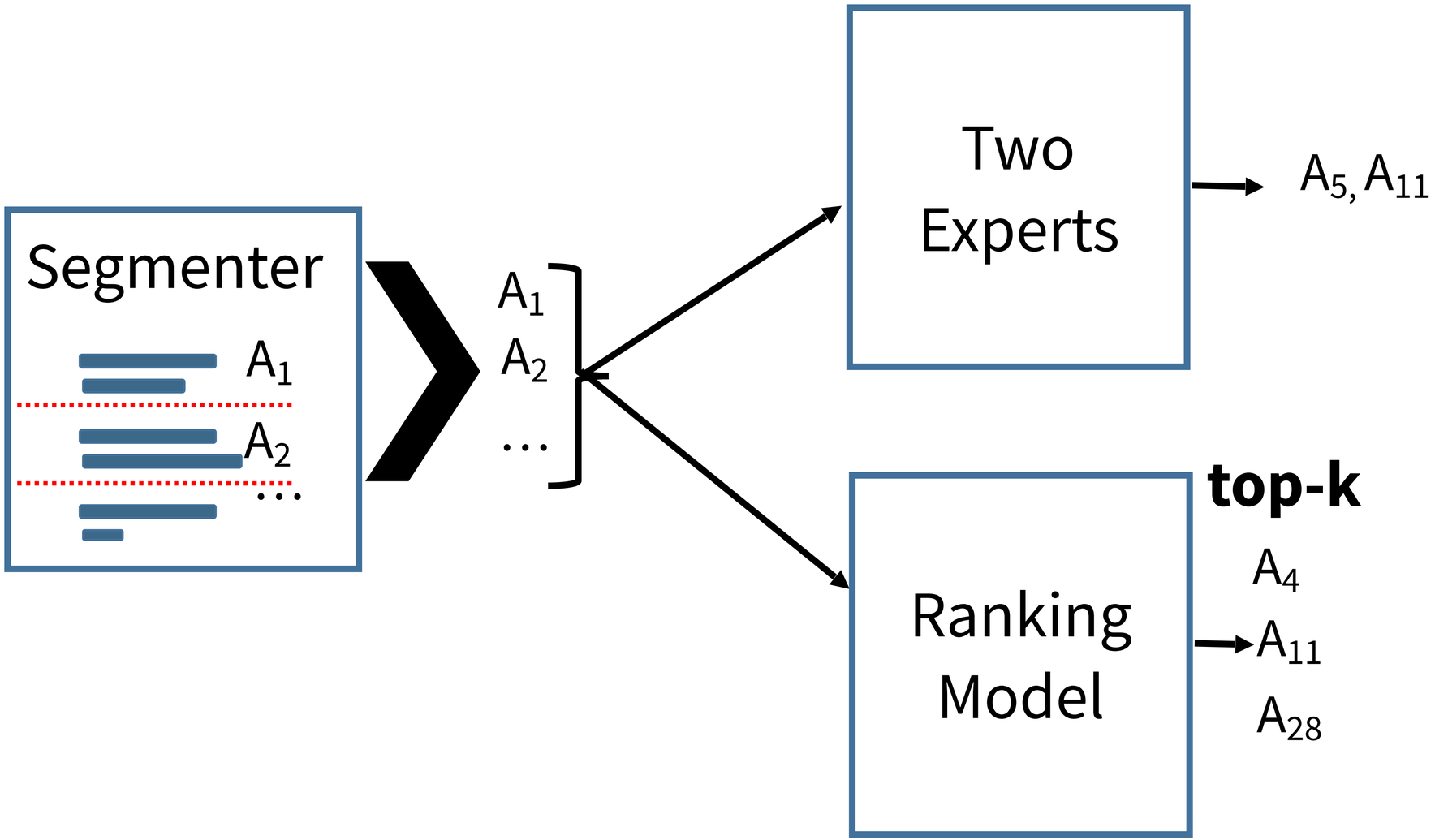
Twitter Dataset

- **120 questions** about **102 companies:**

Evaluation Metrics

- Predictive Accuracy (compared to experts answers)
- User-perceived Utility (how users perceived the answers)

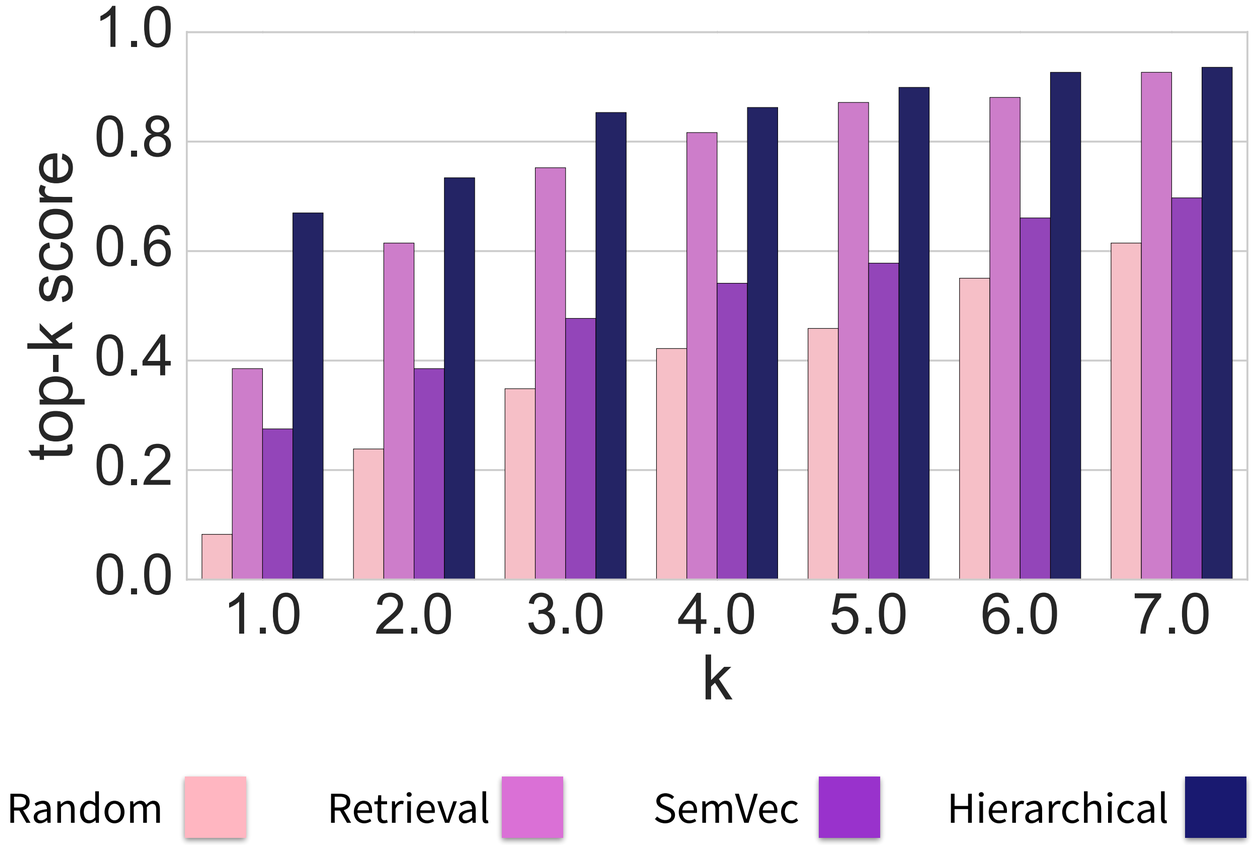
Predictive Accuracy



How many questions have an expert answer in **top-k?**

Predictive Accuracy: top-k score

fraction of Qs with answer among top-k answers



Hierarchical: **82%** of questions had accurate answers in **top 3**

Differences become **less significant** with higher *k*

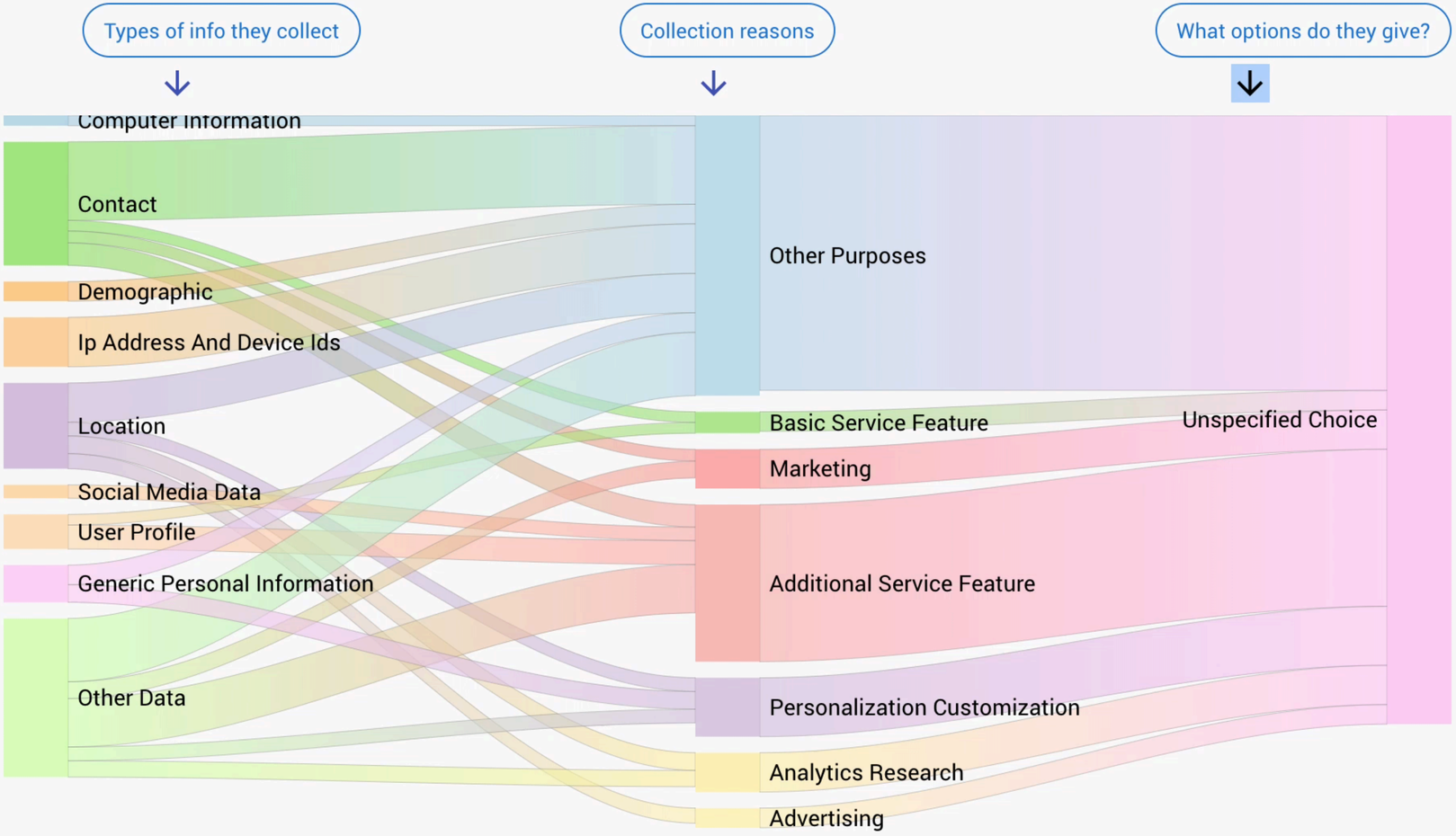
DEMO

tunein.com

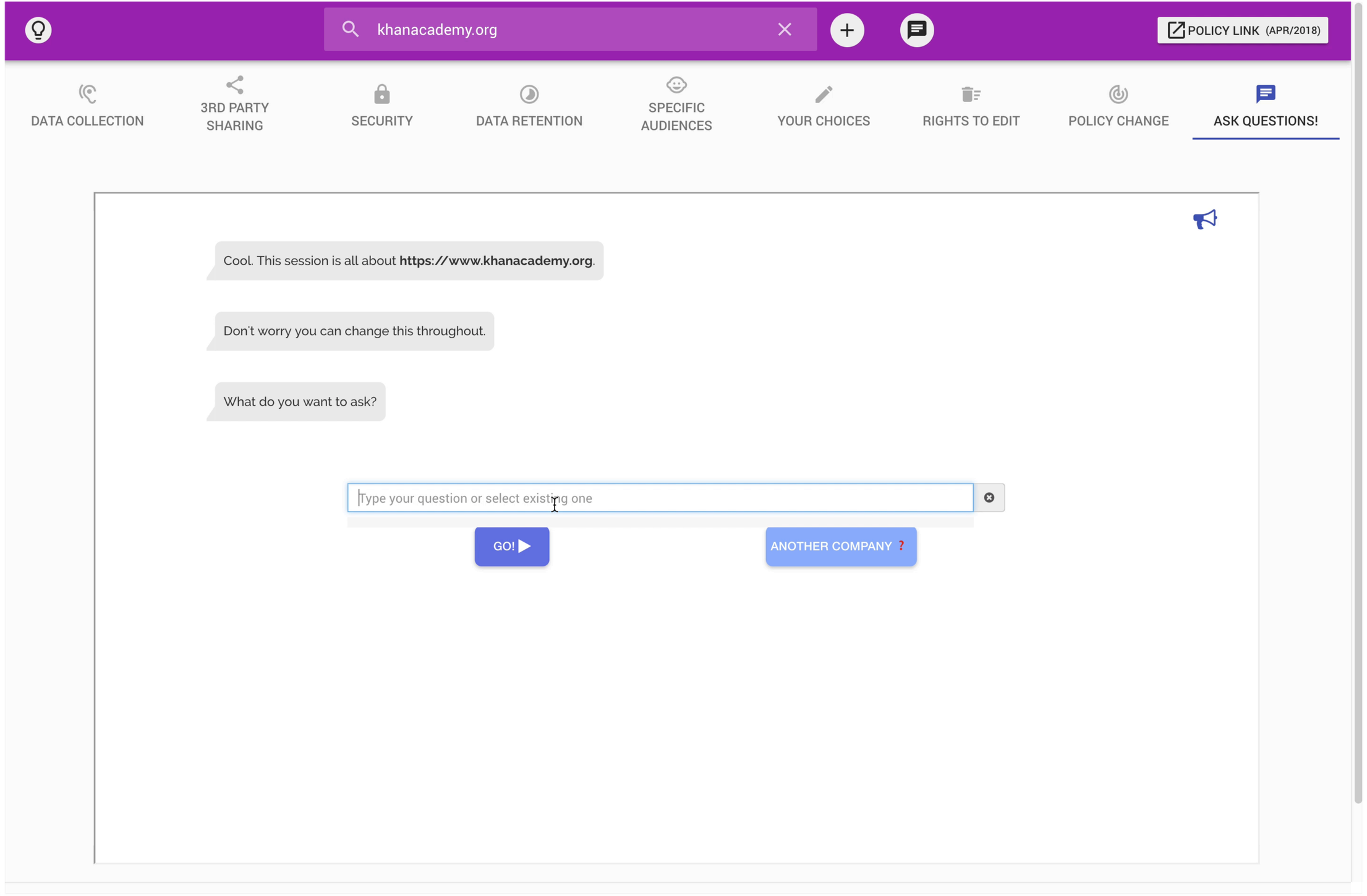
POLICY LINK (DOWNLOADED: 7/AUG/2018)

GOOD/BAD ASPECTS | DATA COLLECTION | 3RD PARTY SHARING | SECURITY | DATA RETENTION | SPECIFIC AUDIENCES | YOUR CHOICES | RIGHTS TO EDIT | POLICY CHANGE | ASK QUESTIONS!

What data does the company gather for its own use?



Choice Links (0)


khanacademy.org
POLICY LINK (APR/2018)
DATA COLLECTION
3RD PARTY SHARING
SECURITY
DATA RETENTION
SPECIFIC AUDIENCES
YOUR CHOICES
RIGHTS TO EDIT
POLICY CHANGE
ASK QUESTIONS!
Cool. This session is all about https://www.khanacademy.org.
Don't worry you can change this throughout.
What do you want to ask?
Type your question or select existing one
GO!
ANOTHER COMPANY ?

Confidence Threshold 0.2

Good

- You can request access and deletion of personal data
- In certain conditions, data is not shared.
- Data is not shared with third parties for advertising purposes.
- The policy states that third parties do not receive personal information.
- The policy offers you clear links to control your data
- Some of the collected data is anonymized or aggregated.

Bad

- Some data might be retained indefinitely.
- Data might be shared in the case of a merger or acquisition.

COMING SOON!

Impact

Users of the app

>35,000

Minutes on our apps

>88,000

Websites analyzed

>21,000

NVIDIA FAST COMPANY Product Hunt WSJ WIRED WORLD ECONOMIC FORUM

Take-aways

- **Polisis**:
  - Unified framework for querying privacy policies
  - Assisting users, regulators, and researchers
  - Two applications:
    - Structured querying: privacy icons generation
    - Unstructured querying: question answering from the privacy policy.

  - **Read more at:**
    - Our paper
    - **WIRED**: Polisis AI Reads Privacy Policies So You Don't Have To
    - **Fast Company**: This Data Viz Tool Explains Privacy Policies You're Too Lazy to Read
    - **WSJ**: Those Privacy Policies Flooding Your Inbox? Print Them Out and They Span a Football Field

pribot.org **hamzaharkous.com**