“I’m not convinced that they don’t collect more than is necessary”: User-Controlled Data Minimization Design in Search Engines

Tanusree Sharma
Assistant Professor, Penn State University | Ph.D., UIUC

Tanusree Sharma
Penn State

Lin Kyi
MPI-SP

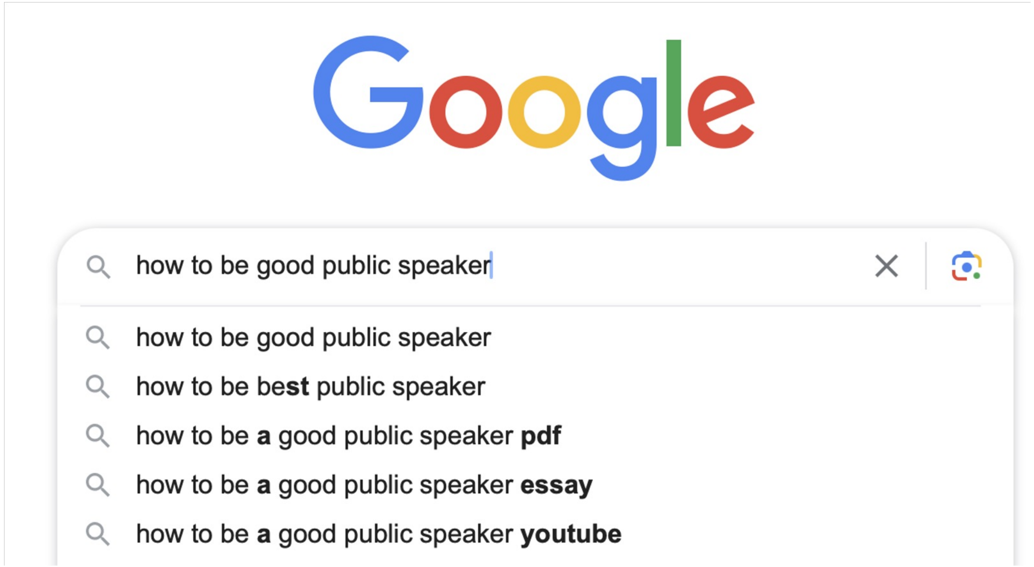
Yang Wang
UIUC

Asia J. Biega
MPI-SP

MAX PLANCK INSTITUTE
FOR SECURITY AND PRIVACY

UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

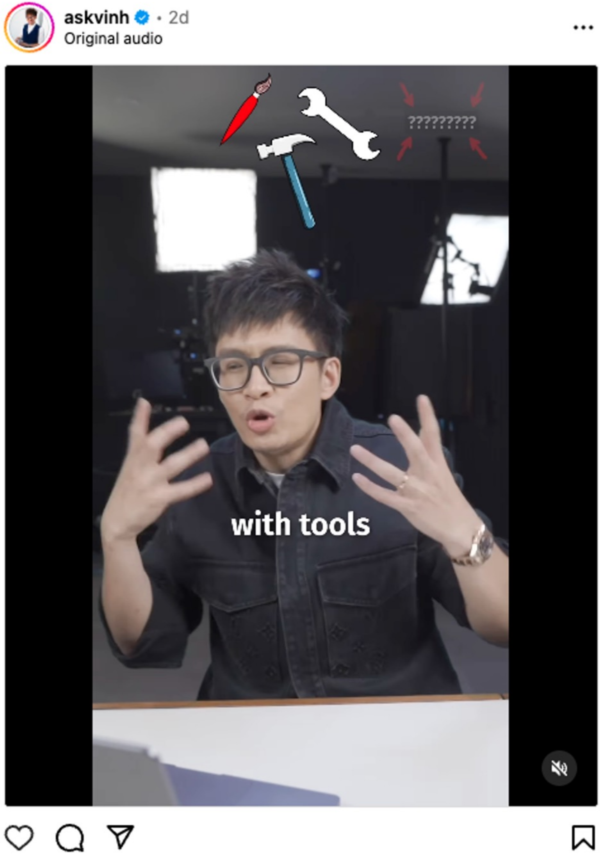
PennState



how to be good public speaker

- how to be good public speaker
- how to be **best** public speaker
- how to be **a** good public speaker **pdf**
- how to be **a** good public speaker **essay**
- how to be **a** good public speaker **youtube**

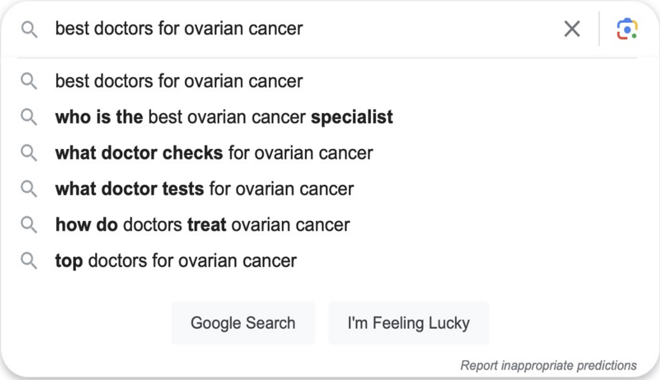
Probably Fine!!!



askvinh · 2d
Original audio

Is it Fine?!

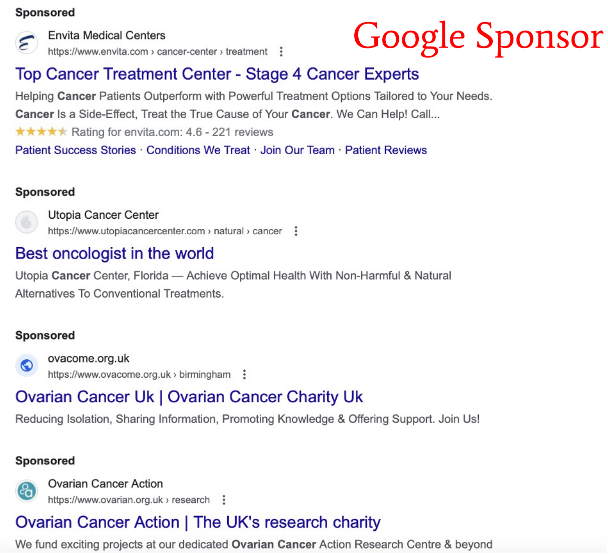
Google



best doctors for ovarian cancer

- best doctors for ovarian cancer
- **who is the** best ovarian cancer **specialist**
- **what doctor checks** for ovarian cancer
- **what doctor tests** for ovarian cancer
- **how do** doctors **treat** ovarian cancer
- **top** doctors for ovarian cancer

Google Sponsor Ads



Sponsored

Envita Medical Centers
https://www.envita.com › cancer-center › treatment

Top Cancer Treatment Center - Stage 4 Cancer Experts

Helping **Cancer** Patients Outperform with Powerful Treatment Options Tailored to Your Needs. **Cancer** Is a Side-Effect, Treat the True Cause of Your **Cancer**. We Can Help! Call...

★★★★★ Rating for envita.com: 4.6 - 221 reviews

Patient Success Stories · Conditions We Treat · Join Our Team · Patient Reviews

Sponsored

Utopia Cancer Center
https://www.utopiacancercenter.com › natural › cancer

Best oncologist in the world

Utopia **Cancer** Center, Florida — Achieve Optimal Health With Non-Harmful & Natural Alternatives To Conventional Treatments.

Sponsored

ovacome.org.uk
https://www.ovacome.org.uk › birmingham

Ovarian Cancer Uk | Ovarian Cancer Charity Uk

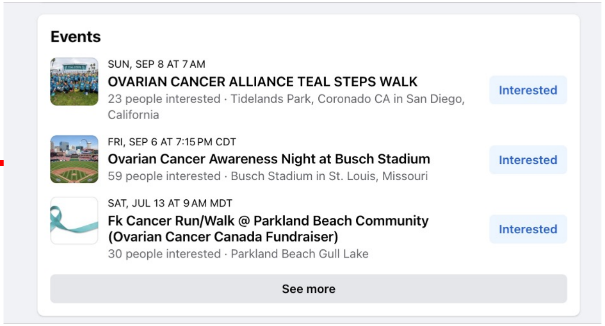
Reducing Isolation, Sharing Information, Promoting Knowledge & Offering Support. Join Us!

Sponsored

Ovarian Cancer Action
https://www.ovarian.org.uk › research

Ovarian Cancer Action | The UK's research charity

We fund exciting projects at our dedicated **Ovarian Cancer** Action Research Centre & beyond



Events

SUN, SEP 8 AT 7 AM
OVARIAN CANCER ALLIANCE TEAL STEPS WALK
23 people interested · Tidelands Park, Coronado CA in San Diego, California
Interested

FRI, SEP 6 AT 7:15 PM CDT
Ovarian Cancer Awareness Night at Busch Stadium
59 people interested · Busch Stadium in St. Louis, Missouri
Interested

SAT, JUL 13 AT 9 AM MDT
Fk Cancer Run/Walk @ Parkland Beach Community (Ovarian Cancer Canada Fundraiser)
30 people interested · Parkland Beach Gull Lake
Interested

See more

Facebook Feed

> *"I live in Hungary and have had a same government for 10-15 years. I live in a non-democratic country. I don't want my health condition to be known it can be dangerous for my benefits when the whole control to the government."*

Data Minimization - A **Legal** Perspective

The EU General Data Protection Regulation (GDPR) outlines the principle of *data minimization* in Article 5(1)(c), requiring that *data should be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed.*

On device policy enforcement “Stakeholders only access the data they need”

Data Minimization During Collection

Designing a Private Logging Pipeline. Mekhola Mukherjee and Thomas Vannet, *Google. Usenix PEPR 2023*

Computational Approach

Performance-Based Data Minimization

Biega et.al Shanmugam et. al.

User Centered Approach

User Interview + Experts Evaluation

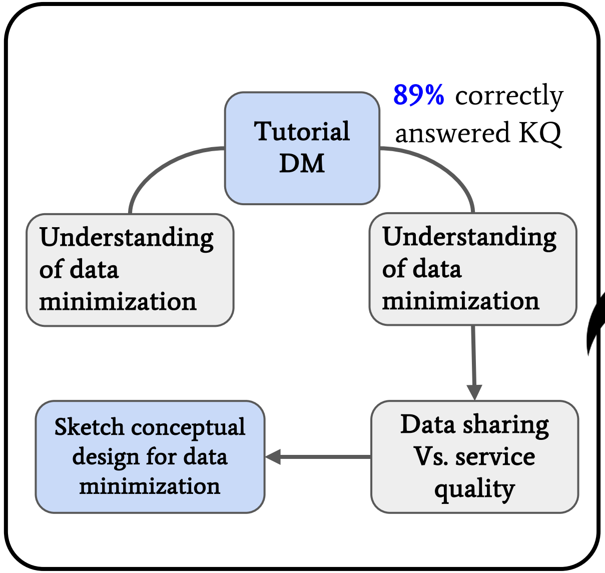
01

RQ1: What factors influence user decisions on what data is **necessary** for search engines to collect in the context of data minimization?

RQ2: How do users think data minimization **ought** to work in search engines? ?

Interview

EU/UK Users (n = 25)



Tutorial DM

89% correctly answered KQ

Understanding of data minimization

Understanding of data minimization

Data sharing Vs. service quality

Sketch conceptual design for data minimization

✓ Decision Making factor

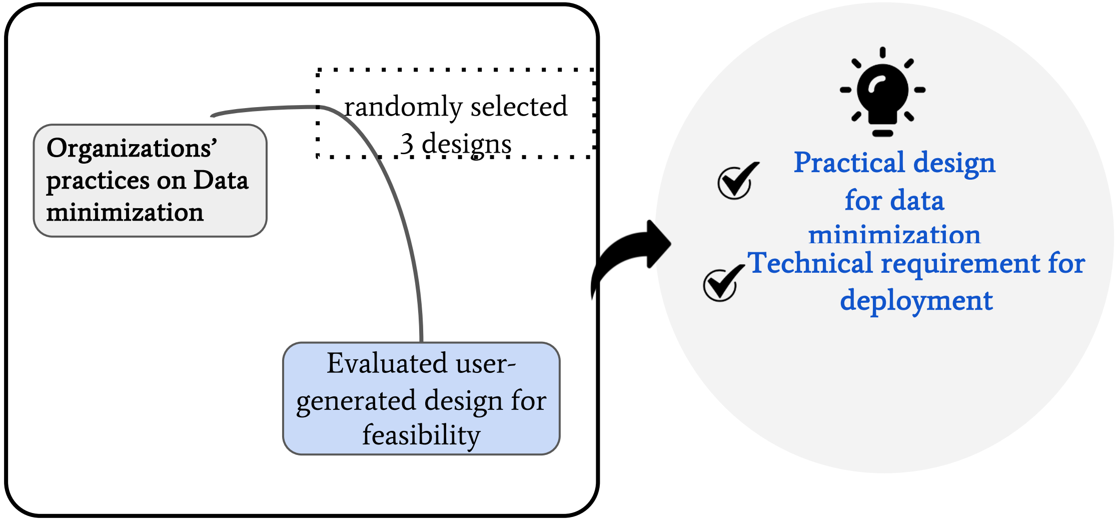
✓ Conceptual Design

02

RQ3: How do experts evaluate the **feasibility** of user generated designs of data minimization in search engines?

Expert Interview

Privacy Experts(n = 10)



Organizations' practices on Data minimization

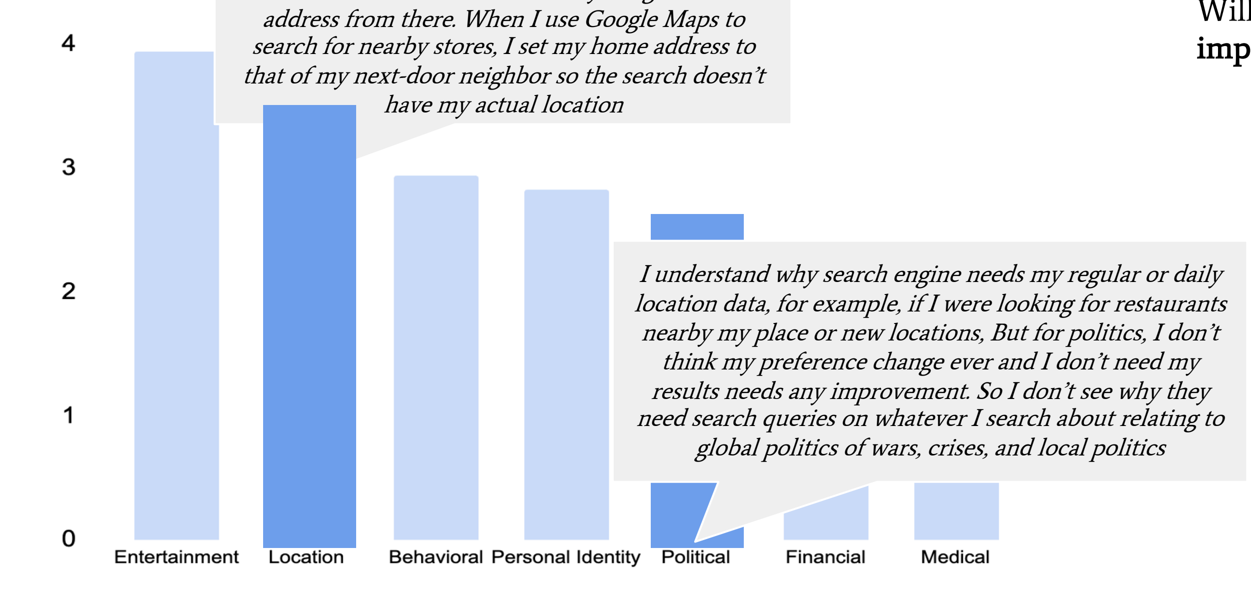
randomly selected 3 designs

Evaluated user-generated design for feasibility

✓ Practical design for data minimization

✓ Technical requirement for deployment

RQ1: Factors Influencing Data Minimization Decisions



I don't think search engine need to store location data to give me good result. If I am searching some coffee shop, it can generate result real time. I am even worried because i searched my daughter school address from there. When I use Google Maps to search for nearby stores, I set my home address to that of my next-door neighbor so the search doesn't have my actual location

I understand why search engine needs my regular or daily location data, for example, if I were looking for restaurants nearby my place or new locations, But for politics, I don't think my preference change ever and I don't need my results needs any improvement. So I don't see why they need search queries on whatever I search about relating to global politics of wars, crises, and local politics

Volume of Search Data needs to provide good service/good search result

Types of your search query data
Willingness to share search data **to improve results for other users**

□ None
□ Few day
□ Few weeks
□ Few months
□ Few years
□ Indefinitely

What is better Result?

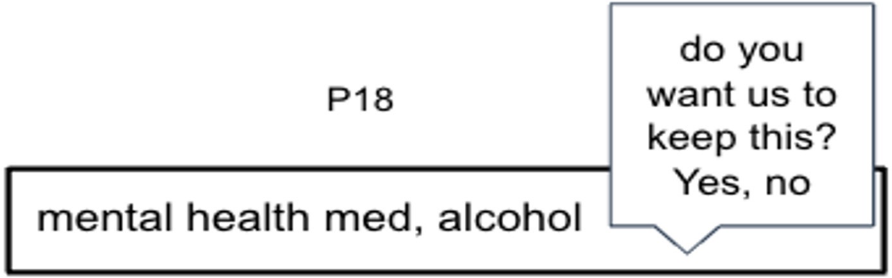
"Top 3-5 result match the search keyword and if see those results are from popular site of what i was search is a better result for me"

RQ2/RQ3: Conceptual Design & Feasibility

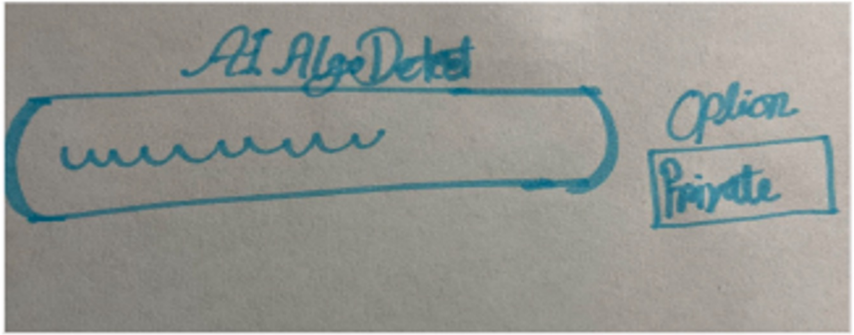
Profile Customization for Data Types and Data Volume
GDPR Alerts for Data Minimization
Separating Searching Sessions for Oneself and Others
Sensitivity-Based Search Customization
Data Donations & Incentives

Conceptual Design & Feasibility- Sensitivity-Based Search

P18


mental health med, alcohol
do you want us to keep this? Yes, no

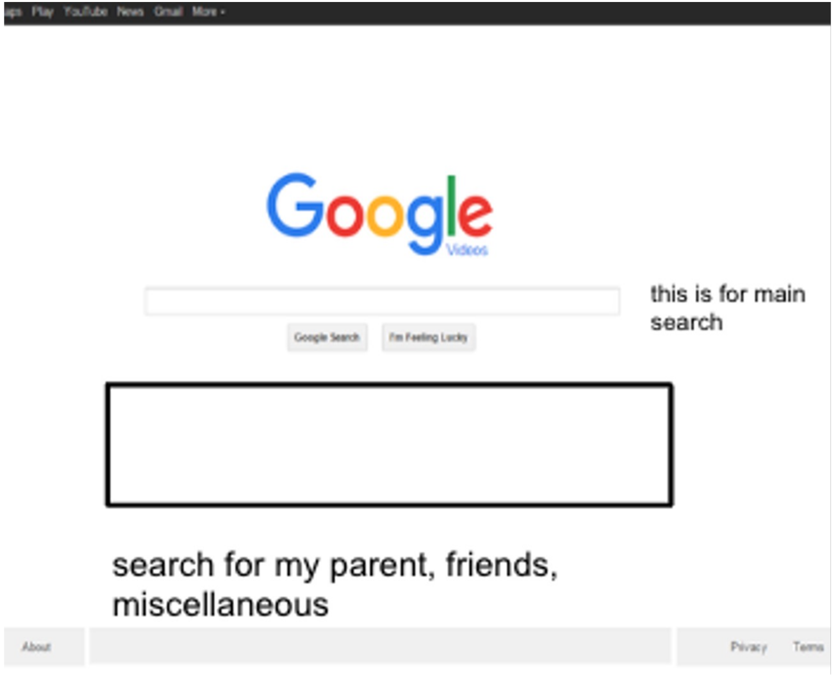
P22



Current engineering or algorithm teams could anonymize or apply access control, adding noise after data collection at the query stage, triggering sensitive search alerts or options. This prevents data from reaching the server, with execution done on the client's search engine. It requires UX design and existing tech stacks in the backend to instantly discard sensitive searches made locally. It's akin to using Trusted Execution Environments (TEEs) for search queries. Cryptography could even be employed for private local execution. From a technical standpoint, this is very feasible."

- -Feasible (Experts in Data Minimization)

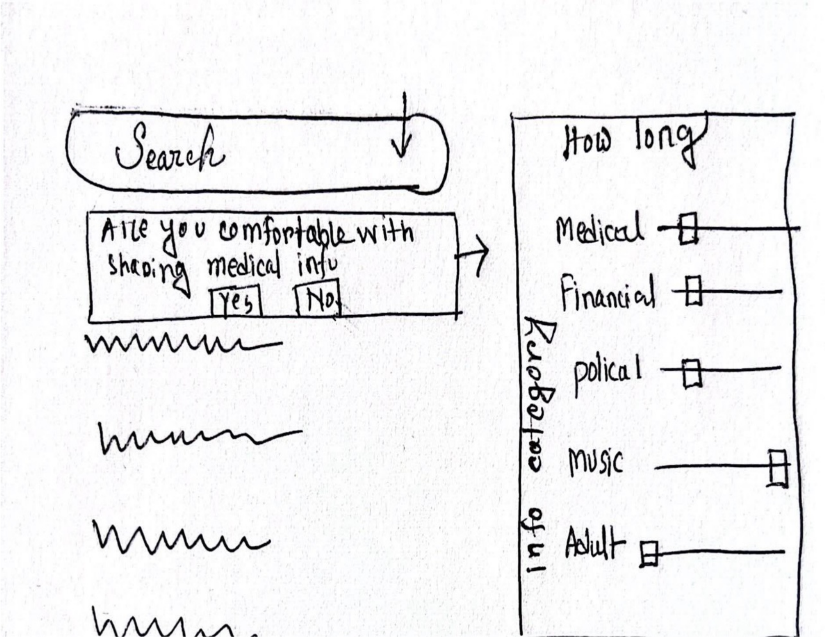
Conceptual Design & Feasibility: Separating Search Session



This is unnecessary data collection. Specifying search queries about another person means re-identification, you are giving another datapoint “other person”, violating privacy of uninvolved parties. Keeping all searches together maintains privacy ***due to the noise****. I wouldn’t recommend this approach.” like your child, essentially you provide the search engine with annotated data to create a shadow profile, detailed footprint of behavior and possibly medical issues of your child under the ’others’ search option. This creates a* ***binary flag from your IP****. This* ***’shadow’ profile,*** *built over let’s say 20 years, could be misused. ”*

—- Not Feasible (Experts in privacy engineering)

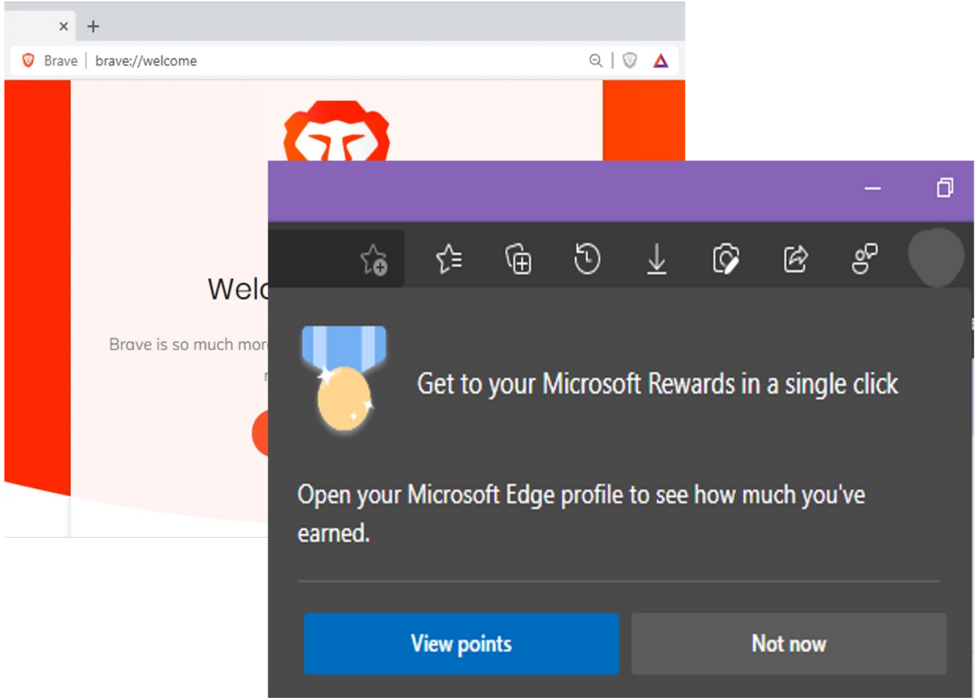
Conceptual Design & Feasibility- Profile Customization



“we ensure secure and efficient data storage. This involves clients’ and their users’ consent, especially in cloud computing. We deal with a variety of rental agreements. At times, we encounter users who prefer to keep their information private, due to the nature of our services spanning multiple countries and continents, and require a certain level of confidentiality for certain types of data. We have a consent-like form design for the product pipelined with db instances to cater to these various requirements efficiently, adding certain noise, adding gibberish for unique identifiers, and setting limits to the data life.”

- -Feasible (Experts in Data Minimization)

Conceptual Design & Feasibility: Data Donations & Incentives



Get to your Microsoft Rewards in a single click

Open your Microsoft Edge profile to see how much you've earned.

Experts: Not Relevant to Data Minimization

Thank you!

Contact:

tfs5747@psu.edu

MAX PLANCK INSTITUTE
FOR SECURITY AND PRIVACY

UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

PennState