Rethinking Data Minimization from a User-Centered Approach

Tanusree Sharma
UIUC

Lin Kyi
MPI-SP

Yang Wang
UIUC

Asia J. Biega
MPI-SP

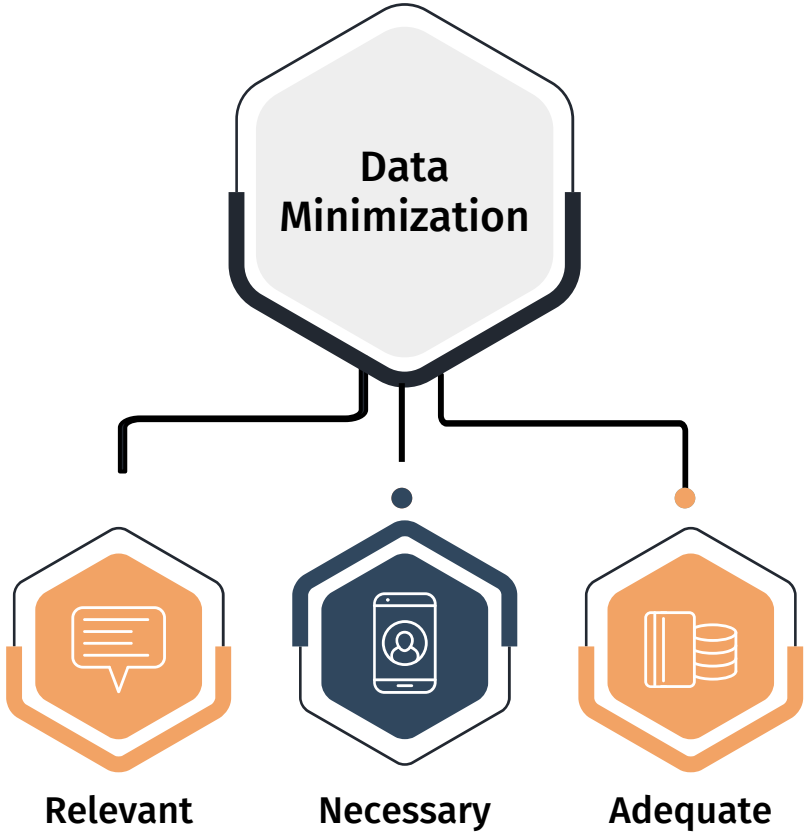
ILLINOIS
School of Information Sciences

Max Planck Institute for
Security and Privacy

PEPR '23

A Legal Perspective

"personal data shall be [...] adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed" — Article 5(1)(c) (GDPR)



System-based approaches

Designing a Private Logging Pipeline

Mekhola Mukherjee and Thomas Vannet, *Google*

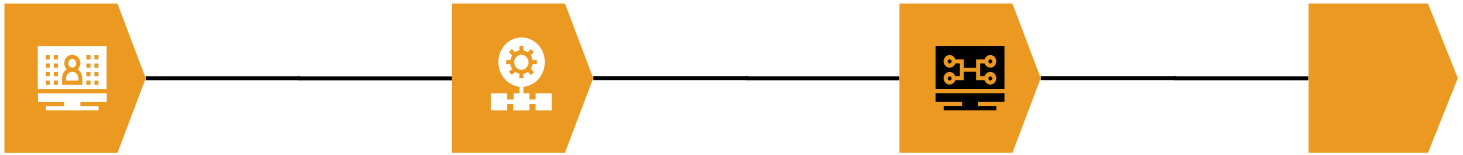
COMPUTATIONAL APPROACH: Literature

Performance-Based Data Minimization

Biega et.al Shanmugam et. al.

USER CENTERED APPROACH

SEMI STRUCTURED INTERVIEW



How Data Minimization Currently Work

Factors Influencing data minimization decision contextually

How Data Minimization ought to work?

Conceptual designing

Current Perception & Willingness

Google Vs Other Search Engine (Edge, Brave, Sapo, Bing, Explorer)

Participants perceived Google as lacking data minimization, while believing other search engines had some measures in place.

Sharing Data to improve service for themselves vs other users

Given a data minimization option, participants are more willing to share data to benefit users with similar interests than to improve their own search results.

Factors Influencing data minimization decision contextually

Required Volume of Data

Comprehension regarding the amount of data required for good results from search engines. They highlighted distinct search data attributes: (a) quantity of data in temporal proximity (past/recent)

How Much Data ~ for How Much Improvement

Decision making factor regarding the minimization of data collection and usage for intended service. Users considered the trade-offs between the level of information and the extent of improvement in the search result that could be achieved

Conceptual Designs of User-Driven Data Minimization in Search Engines

Sensitivity-Based Search

Selection & Customization Based on Data Type and Volume

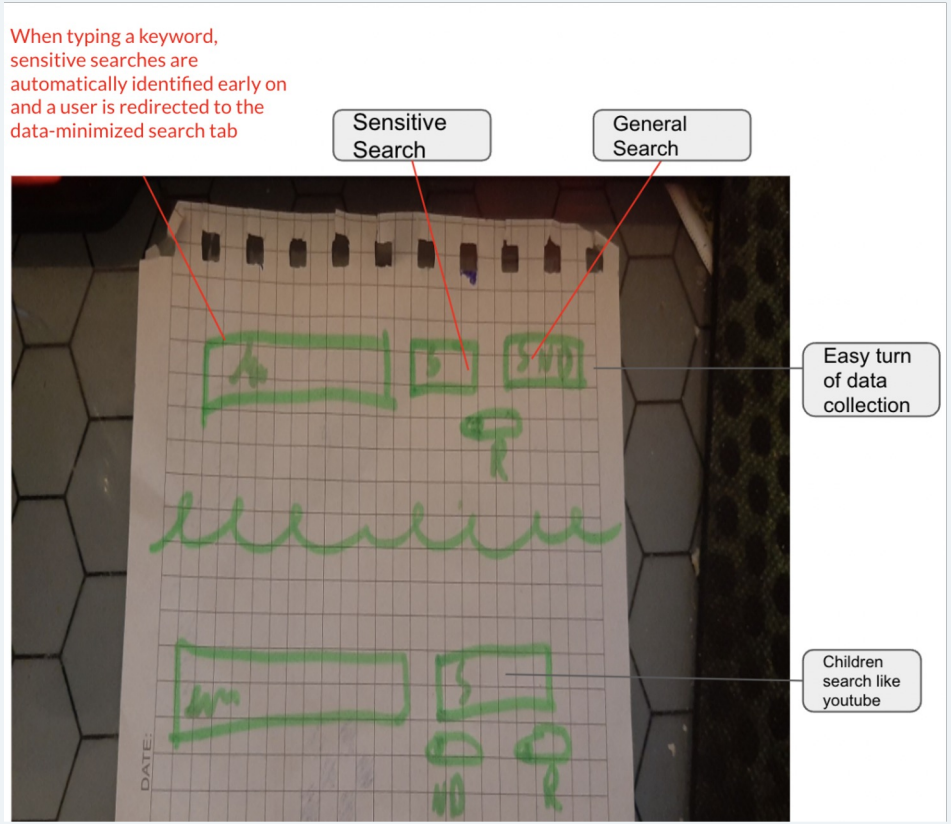
Visualization of Data Necessity & Relevancy for Purpose

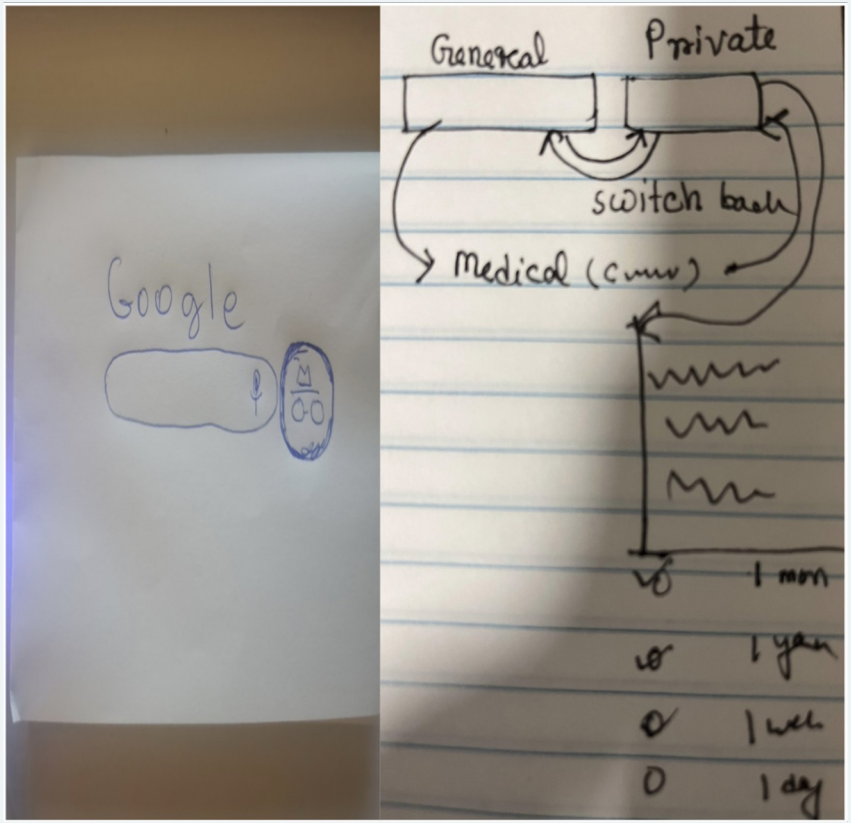
Option to Separate if Searching for Themselves or Others

Proper Consent Mechanisms

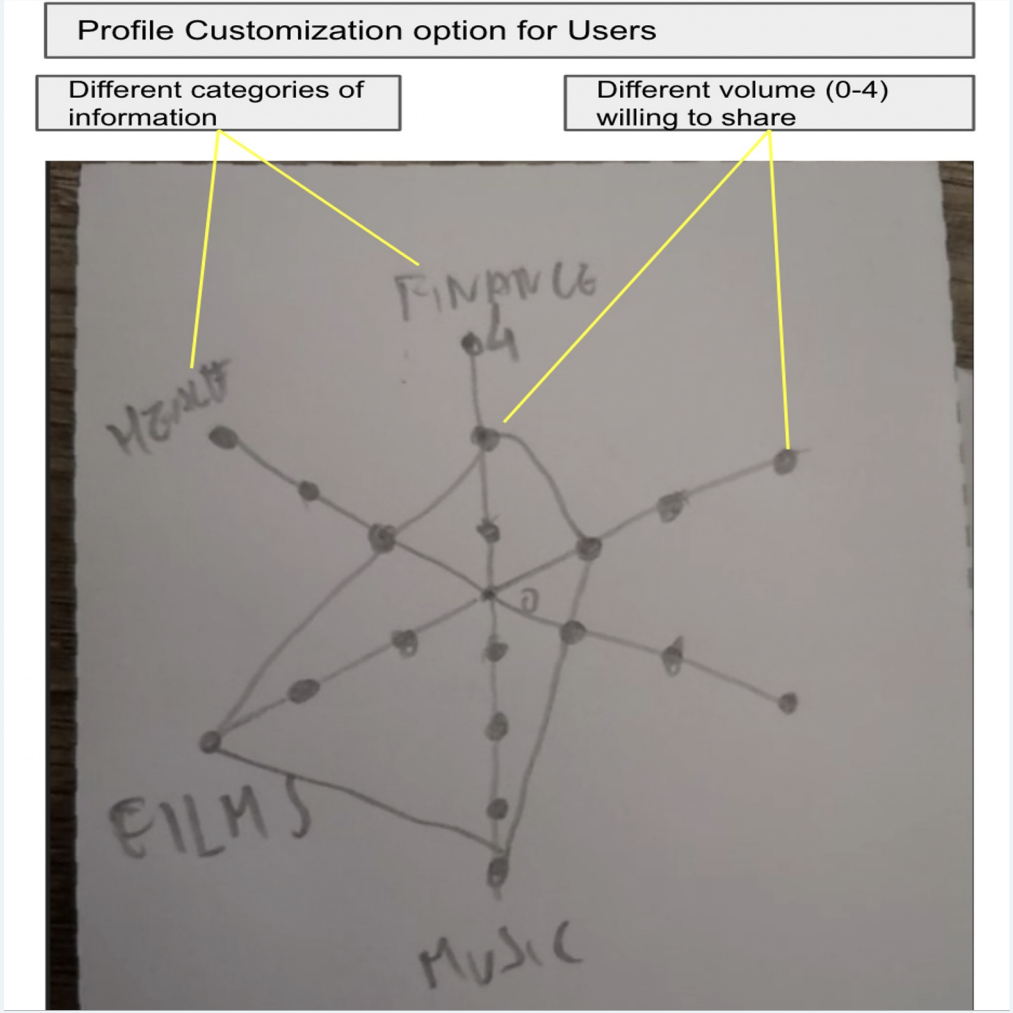
Sensitivity-Based Search

When typing a keyword, sensitive searches are automatically identified early on and a user is redirected to the data-minimized search tab

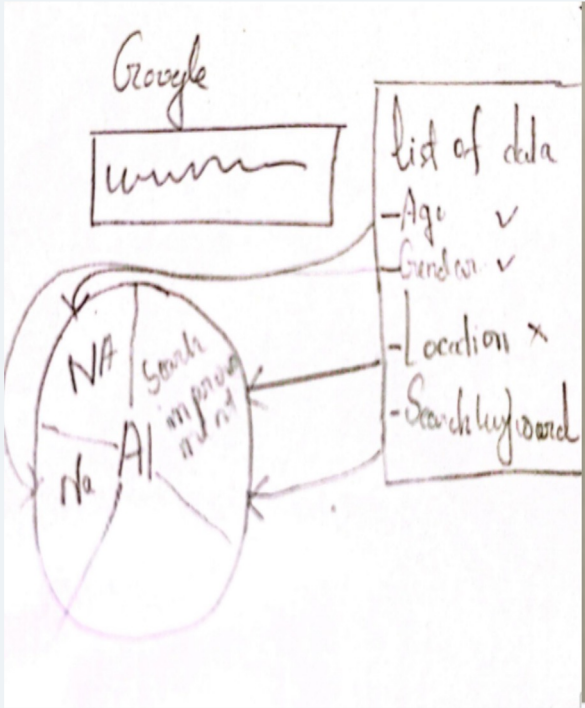




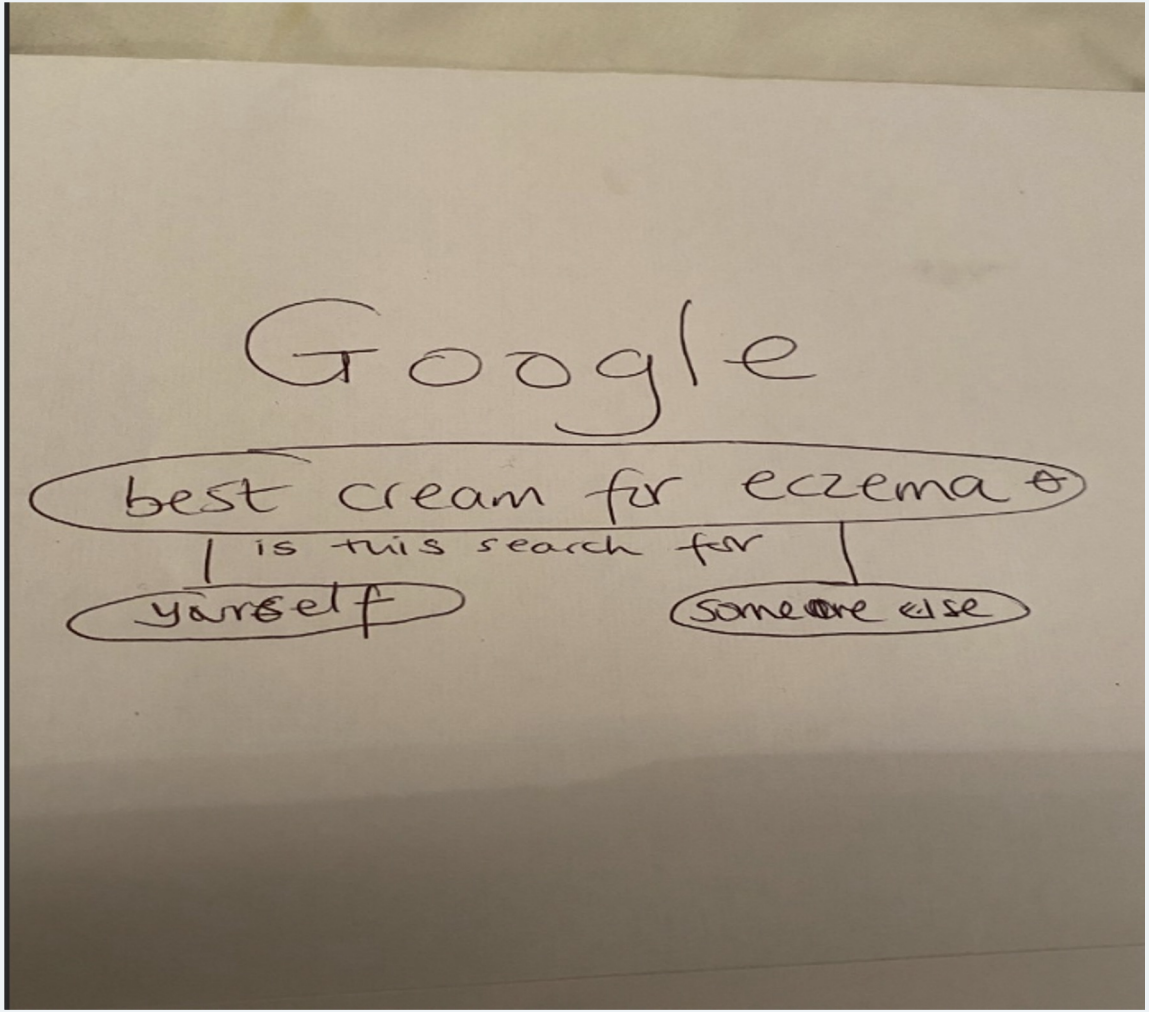
Fine Grained Selection & Customization Based on Data Type and Volume



Visualization of Data Necessity & Relevancy for Purpose



Option to Separate if Searching for Themselves or Others



How do we validate if this designs are feasible and Practical?

DATA MINIMIZATION AS RISK MANAGEMENT TOOL

Existing Risk Framework

Risk Identification

- Data inventory
- Application Risk Profile

Risk Assessment

- Sensitivity score of Data
- Organizational impact score
- Control strength

Review Decision

- Assessment of residual risk
- Qualitative input from experts

End Users

Values

In form of perception and conceptual design

Factors

That influence their decision making

reference: https://www.nist.gov/privacy-framework

How to engage people (scale) in this process when using data driven services?
How do we keep up with evolved preferences of users overtime?

Reimagine how we would like users to interact with technology

Configurations of power

Politicizing technology (e.g., Deliberative decision making)

Outcome/Quality of community decisions

Image source: https://www.flaticon.com/

Thank you!

Contact:
tsharma6@illinois.edu

@Tanusree_Sharma

Key Takeaways

- Data minimization as risk management tooling
- Users conceptual design can be rich data, which need further well informed expert reviews
- Future work:
  - Scaling of end users engagement is challenging
    - Possible way is utilizing already existing protocols from Decentralized governance

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Reference: https://slidesgo.com/