SHOULD I TRUST IT WITH MY DATA?

CAPABILITIES, LIMITS, AND PERSPECTIVES OF AI TECHNOLOGIES FOR PRIVACY

Victor Morel

SPAIML2025

CHALMERS
UNIVERSITY OF TECHNOLOGY

THE AI WAVE

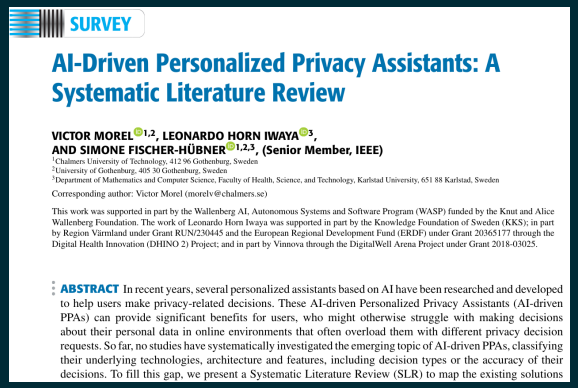
AI IS EVERYWHERE, AND IT DISRUPTS EVERYTHING

BUT WHAT DOES AI DO TO PRIVACY?

- Does it enhance it or undermine it?
- Is automation always desired and always desirable
- Can we even legally introduce AI in all aspects of our lives?
- What are the considerations one must take to integrate AI in privacy-sensitive systems?

OUTLINE

AI for privacy



SURVEY

AI-Driven Personalized Privacy Assistants: A Systematic Literature Review

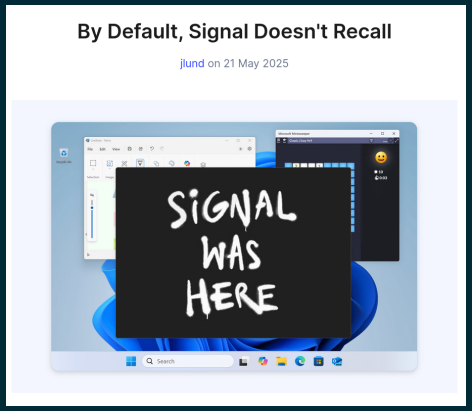
VICTOR MOREL[1,2], LEONARDO HORN IWAYA[3], AND SIMONE FISCHER-HÜBNER[1,2,3], (Senior Member, IEEE)
[1]Chalmers University of Technology, 412 96 Gothenburg, Sweden
[2]University of Gothenburg, 405 30 Gothenburg, Sweden
[3]Department of Mathematics and Computer Science, Faculty of Health, Science, and Technology, Karlstad University, 651 88 Karlstad, Sweden
Corresponding author: Victor Morel (morelv@chalmers.se)

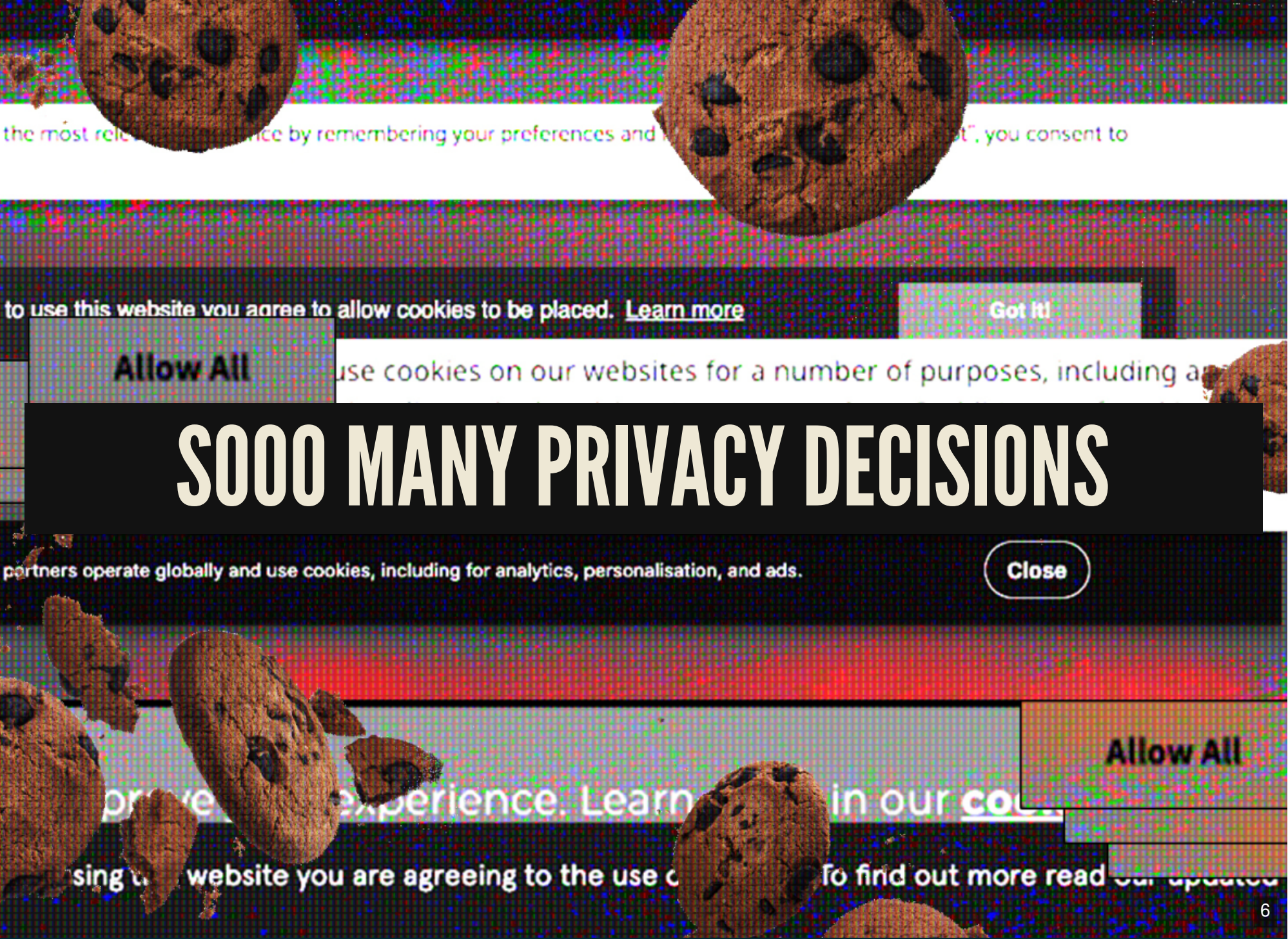
This work was supported in part by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. The work of Leonardo Horn Iwaya was supported in part by the Knowledge Foundation of Sweden (KKS); in part by Region Värmland under Grant RUN/230445 and the European Regional Development Fund (ERDF) under Grant 20365177 through the Digital Health Innovation (DHINO 2) Project; and in part by Vinnova through the DigitalWell Arena Project under Grant 2018-03025.

ABSTRACT In recent years, several personalized assistants based on AI have been researched and developed to help users make privacy-related decisions. These AI-driven Personalized Privacy Assistants (AI-driven PPAs) can provide significant benefits for users, who might otherwise struggle with making decisions about their personal data in online environments that often overload them with different privacy decision requests. So far, no studies have systematically investigated the emerging topic of AI-driven PPAs, classifying their underlying technologies, architecture and features, including decision types or the accuracy of their decisions. To fill this gap, we present a Systematic Literature Review (SLR) to map the existing solutions

Privacy for AI?

AI against privacy




SOOO MANY PRIVACY DECISIONS
Allow All
Got it!
Close
Allow All

PEOPLE HAVE BEEN BUILDING PRIVACY ASSISTANTS TO ADDRESS THE ISSUE

And many of these assistants now embed AI

Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions

Bin Liu,* Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi
Shikun Zhang, Norman Sadeh,* Alessandro Acquisti, Yuvraj Agarwal
Carnegie Mellon University
Pittsburgh, PA, USA
{bliu1, manderse, fschaub, hazim, shikunz, sadeh, yuvraj.agarwal}@cs.cmu.edu
acquisti@andrew.cmu.edu

ABSTRACT

Modern smartphone platforms have millions of apps, many of which request permissions to access private data and resources, like user accounts or location. While these smartphone platforms provide varying degrees of control over these permissions, the sheer number of decisions that users are expected to manage has been shown to be unrealistically high. Prior research has shown that users are often unaware of, if not uncomfortable with, many of their permission settings. Prior work also suggests that it is theoretically possible to predict many of the privacy settings a user would want by asking the user a small number of questions. However, this approach has neither been operationalized nor evaluated with actual users before. We report on a field study (n=72) in which we implemented and evaluated a Personalized Privacy Assistant (PPA) with participants using their own Android devices. The results of our study are encouraging. We find that 78.7% of the recommendations made by the PPA were adopted by users. Following initial recommendations on permission settings, participants were motivated to further review and modify their settings with daily "privacy nudges." Despite showing substantial engagement with these nudges, participants only changed 5.1% of the settings previously adopted based on the PPA's recommendations. The PPA and its recommendations were perceived as useful and usable. We discuss the implications of our results for mobile permission management and the design of personalized privacy assistant solutions.

While the Android and iOS platforms both rely on permission-based mechanisms and allow users to control access to sensitive data and functionality, the end result is an unwieldy number of app-permission decisions that users are expected to make. Estimates indicate that users, on average, have to make over one hundred permission decisions (95 installed apps on average per user [48]; 5 permissions on average per app [37]). Prior work has shown that users are often unaware of – if not uncomfortable with – many of the permissions they have ostensibly consented to at some point (e.g., [6, 8, 16, 17, 21, 24]).

To help overcome the burden associated with managing such a large number of decisions, prior research suggests that – despite the diversity of users' privacy preferences – it is theoretically possible to predict many of a user's permission settings by asking the user a small number of questions [28, 29]. These approaches suggest that, using machine learning, it may be possible to reduce user burden when it comes to configuring mobile app permission settings. However, this approach has not been fully operationalized so far.

We propose a practical solution that operationalizes privacy preference modeling in a personalized privacy assistant (PPA) by (1) developing privacy profiles for users, (2) determining which of these profiles is the best match for a given user, and (3) configuring many of the user's permissions based on the selected profile. This paper is the first to report on the implementation and field evaluation of a personalized privacy assistant (PPA) for mobile app permissions.

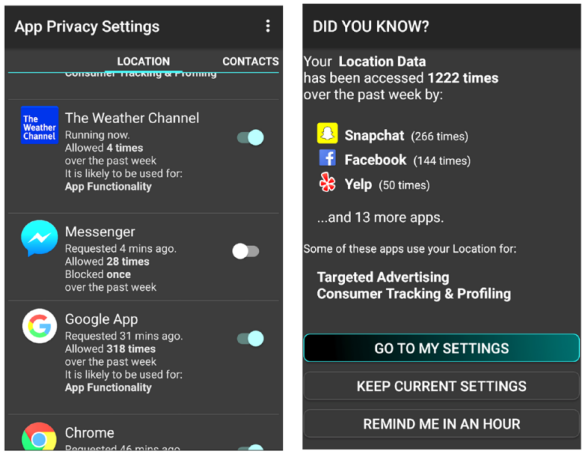


Figure 1: Permission manager (*left*) and a daily privacy nudge (*right*), which include the access frequency and purpose information.

BUT WHAT IS THE LITERATURE ACTUALLY SAYING?

- How many papers were published?
- What are the capabilities of these AI-driven PPAs?
- How is AI integrated?
- Do they respect legal requirements?
- Which decisions are concerned?

👉 SYSTEMATIC LITERATURE REVIEW

SURVEY

AI-Driven Personalized Privacy Assistants: A Systematic Literature Review

VICTOR MOREL[1,2], **LEONARDO HORN IWAYA**[3], **AND SIMONE FISCHER-HÜBNER**[1,2,3], **(Senior Member, IEEE)**

[1]Chalmers University of Technology, 412 96 Gothenburg, Sweden
[2]University of Gothenburg, 405 30 Gothenburg, Sweden
[3]Department of Mathematics and Computer Science, Faculty of Health, Science, and Technology, Karlstad University, 651 88 Karlstad, Sweden

Corresponding author: Victor Morel (morelv@chalmers.se)

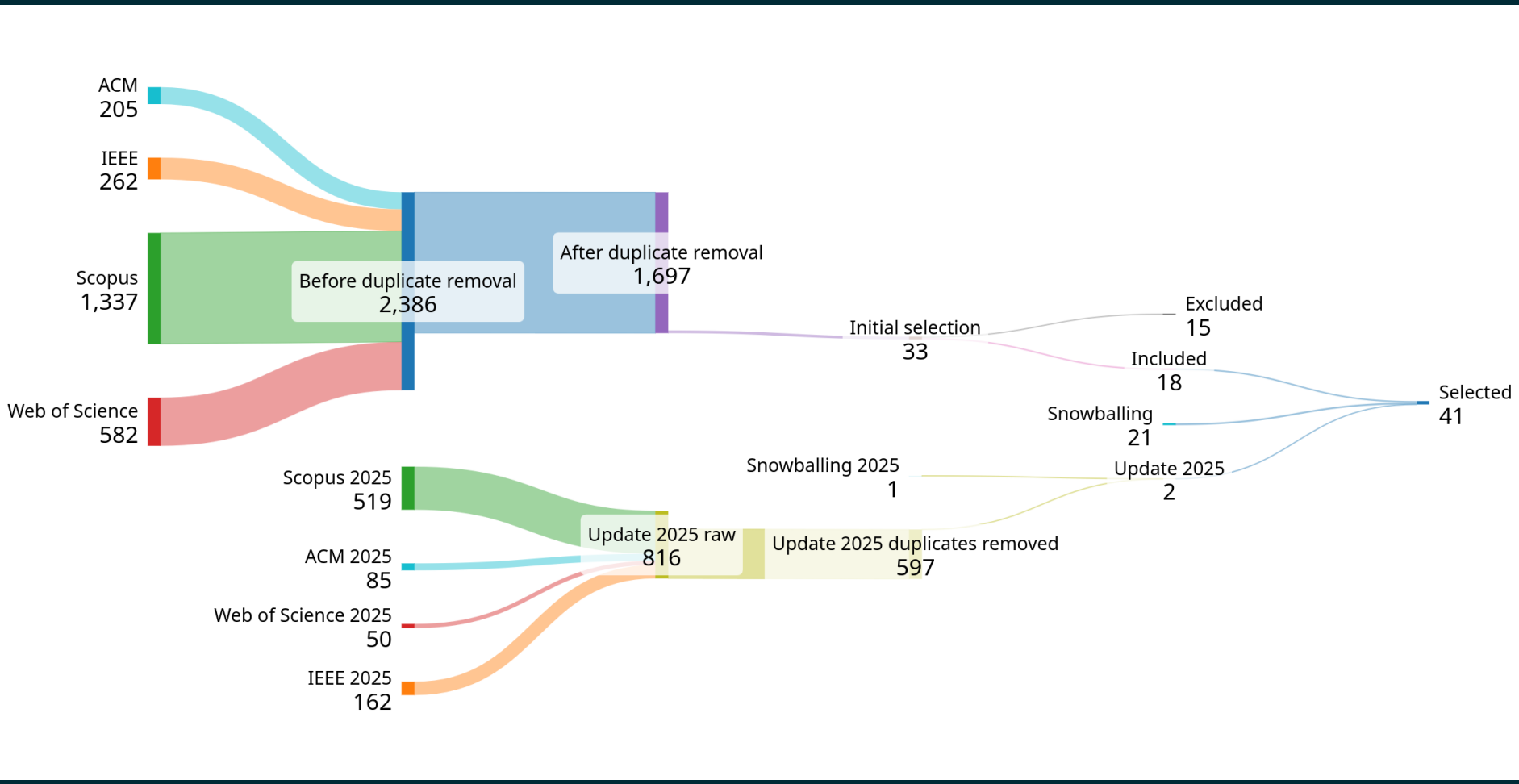
This work was supported in part by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Ali Wallenberg Foundation. The work of Leonardo Horn Iwaya was supported in part by the Knowledge Foundation of Sweden (KKS); in pa by Region Värmland under Grant RUN/230445 and the European Regional Development Fund (ERDF) under Grant 20365177 through t Digital Health Innovation (DHINO 2) Project; and in part by Vinnova through the DigitalWell Arena Project under Grant 2018-03025.

ABSTRACT In recent years, several personalized assistants based on AI have been researched and develo to help users make privacy-related decisions. These AI-driven Personalized Privacy Assistants (AI-dri PPAs) can provide significant benefits for users, who might otherwise struggle with making decisi about their personal data in online environments that often overload them with different privacy decis requests. So far, no studies have systematically investigated the emerging topic of AI-driven PPAs, classify their underlying technologies, architecture and features, including decision types or the accuracy of t decisions. To fill this gap, we present a Systematic Literature Review (SLR) to map the existing soluti

OUR METHODOLOGY

```
1   Search Query = {
2   (privacy OR "data protection") AND
3   (assistant* OR agent*) AND
4   ("artificial intelligence" OR "machine*learning"
5   OR intelligent OR automat* OR personali*ed)
6 }
```

SELECTION PROCESS



Classification overview


AI-driven PPA
Source of data
AI used
Behavioral
Content of data
Transparent
Non-inherently transparent
Past decisions
Current settings
Clustering
Rule-based
Classification
Reinforcement
Attitudinal
Data type
Logic-based
LLM
Metadata
Context
Type of decision
System context
Preferences
Permissions
IoT
Mobile apps
Data sharing
Social media
Cloud
Intelligent retail stores
Empirical assessment
User control over decisions
Dataset
Accuracy measurement
Informed
Semi-automated
User study
Specific
Revoke
Architecture
Local computation
Remote computation
Federated learning

TYPE OF DECISION

Permissions

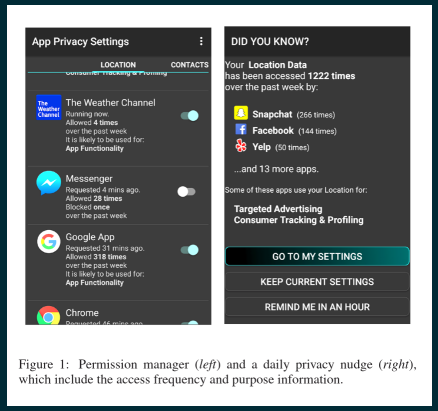
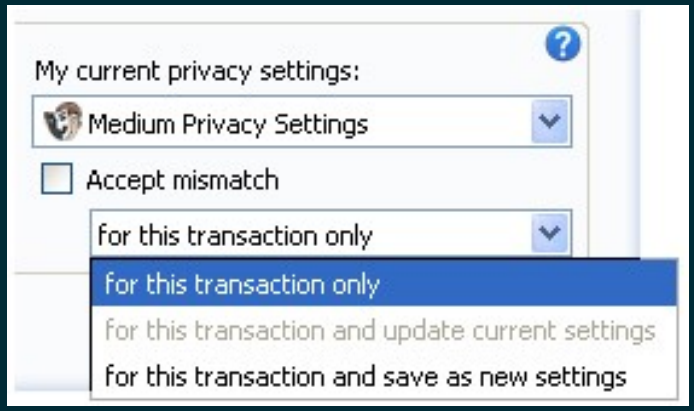
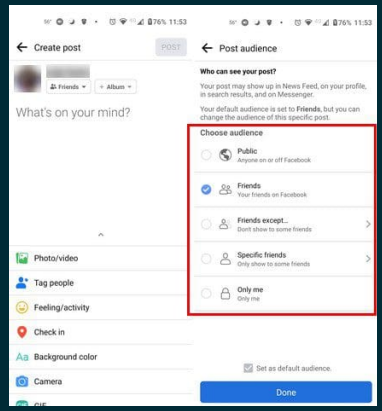


Figure 1: Permission manager (*left*) and a daily privacy nudge (*right*), which include the access frequency and purpose information.

Preferences



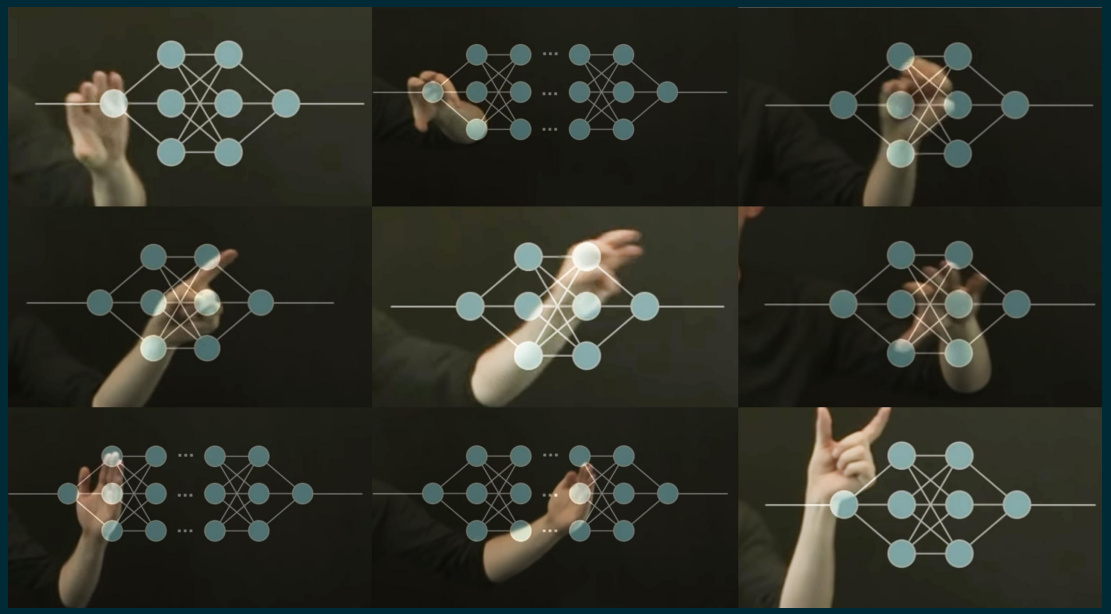
Data sharing



AI USED

Transparent

Non-intrinsically Transparent



SOURCE OF DATA

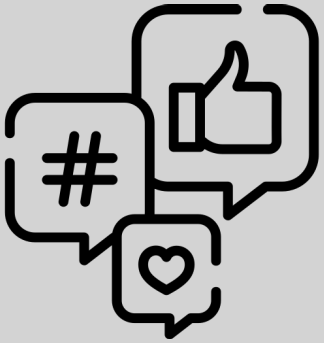
- Context
- Attitudinal data
- Behavioral data
- Metadata
- Data type
- Content of data

SYSTEM CONTEXT

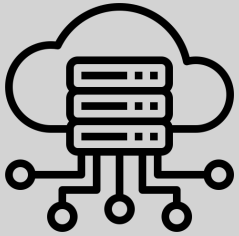
Mobile apps

IoT

Social Media



Cloud



Intelligent retail stores

LESSONS LEARNED

EVALUATION OF THESE PPAS IS POOR

1. THE EVALUATIONS OF AI-DRIVEN PPAS ARE NOT BASED ON THE SAME OR COMPARABLE ACCURACY METRICS OR MEASUREMENTS.

2. OUR DATA SHOWS A LACK OF USER STUDY EVALUATIONS.

👉 STANDARD MEASUREMENT OF ACCURACY AND EVALUATION, NOTABLY THROUGH USER STUDIES.

DECISIONS ARE USUALLY NOT PROPERLY EXPLAINED NOR EVEN EXPLAINABLE

1. ONLY ONE OF THE SURVEYED PAPERS EXPLICITLY ADDRESSES EXPLAINABILITY.

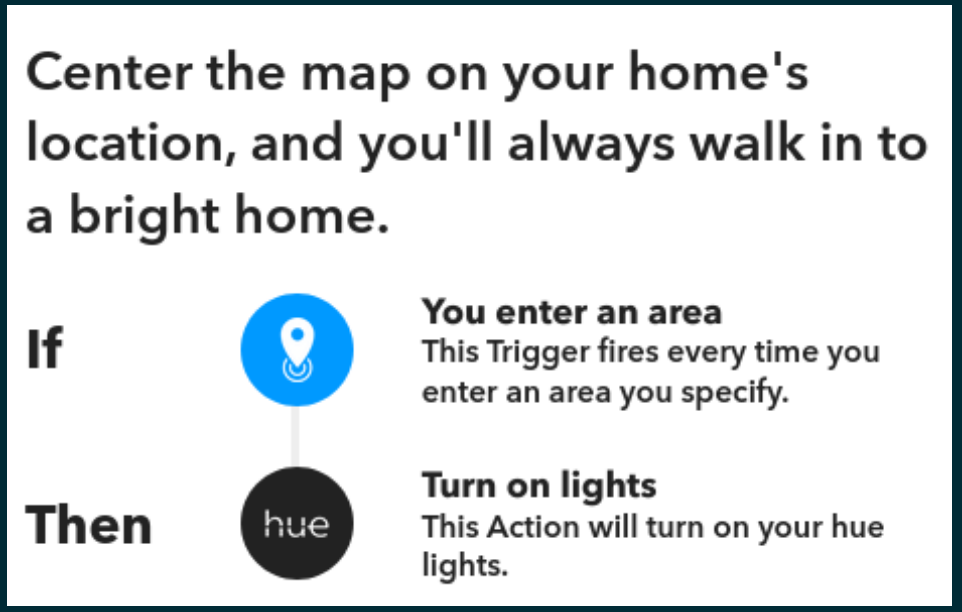
2. WHILE TRANSPARENCY CAN BE LEGALLY REQUIRED, AND FOSTERS TRUST ANYWAY.

👉 INCLUDE EXPLANATION MECHANISMS, POST-HOC IF NEEDED.

SOME SYSTEM CONTEXTS ARE MISSING

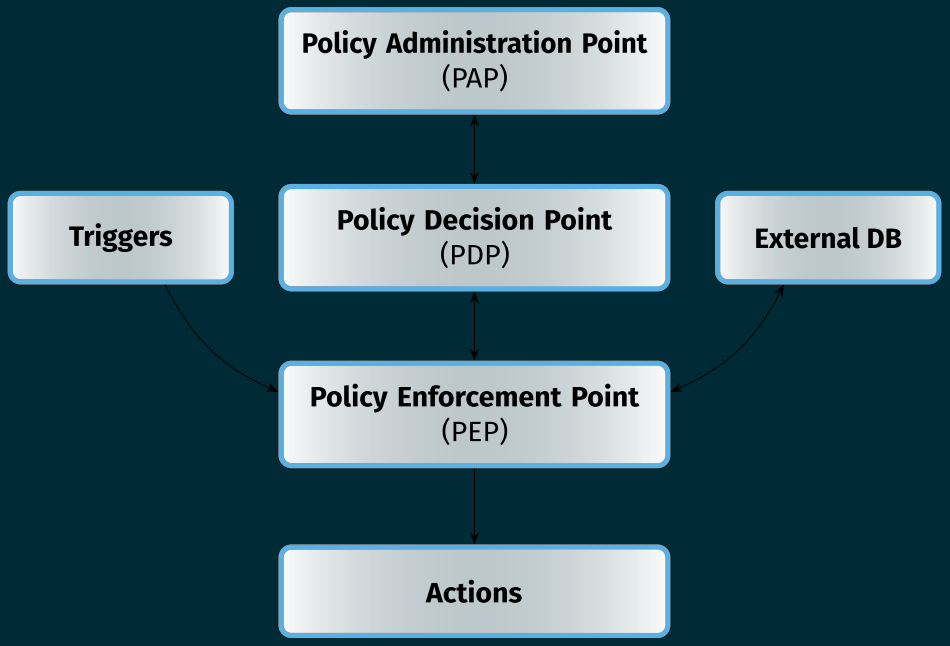
1. NO PERSONALIZED ASSISTANT FOR COOKIE BANNERS.

2. NO WORK ON EMERGING TECHNOLOGY SUCH AS TAPS.



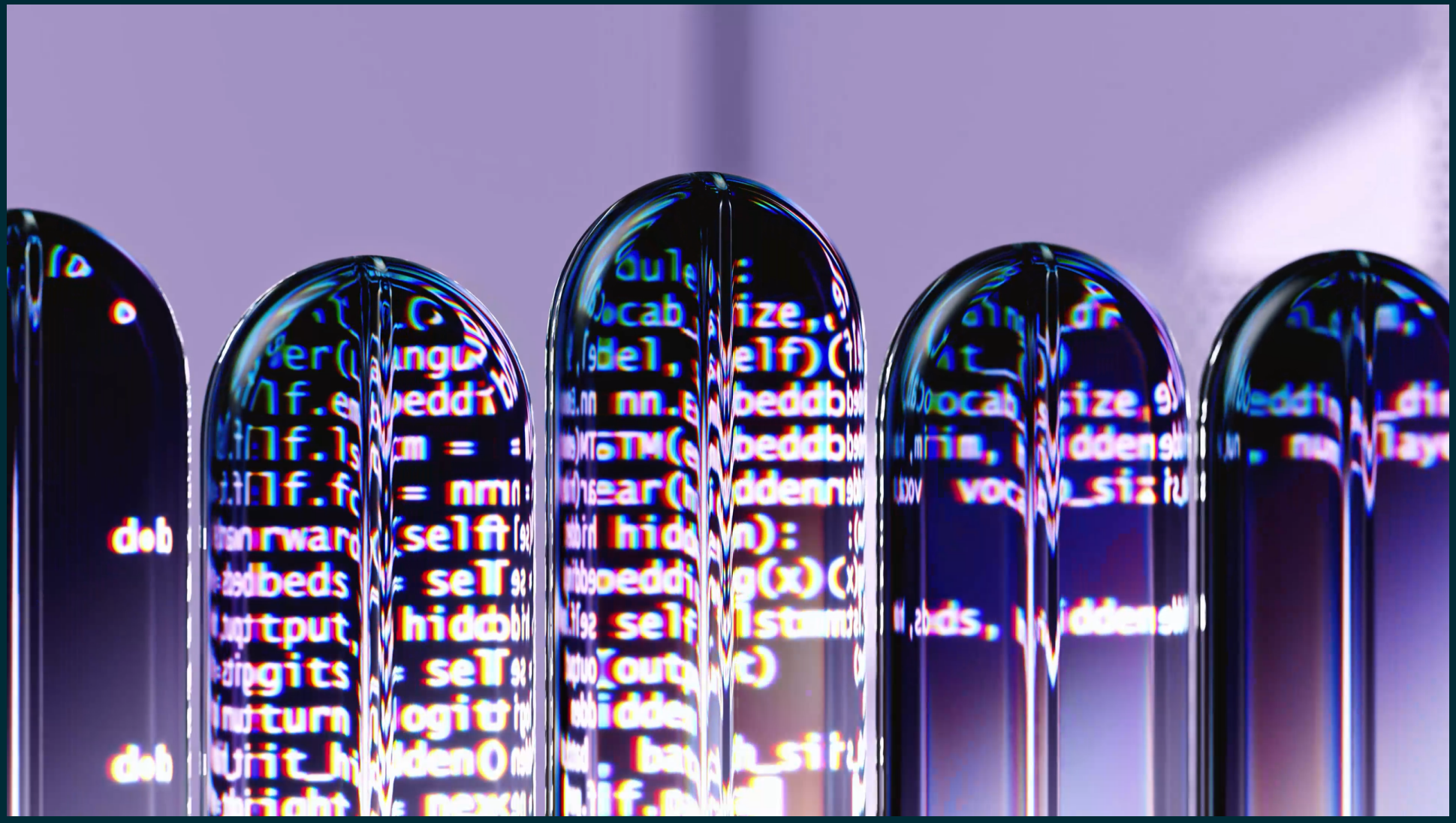
FUTURE WORK

WE ARE BUILDING AN AI-DRIVEN PPA FOR TRIGGER-ACTION PLATFORMS

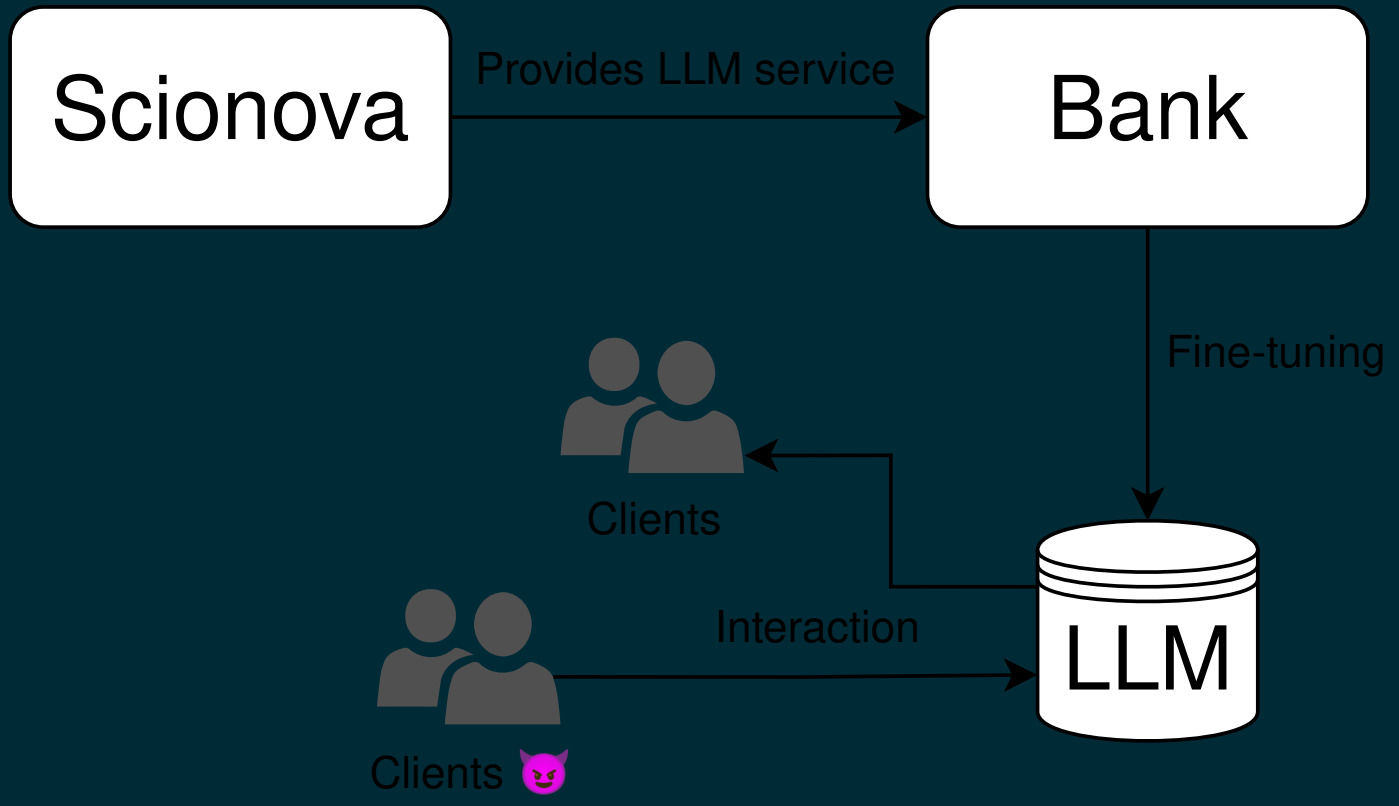


- Functioning prototype based on privacy profiles
- Ongoing work on the UIs
- Upcoming integration of ML models for better personalization

WHAT ABOUT LLM PRIVACY?



INDUSTRIAL MASTER THESIS STUDYING PRIVACY/EFFICIENCY TRADEOFFS



PRIVACY ATTACKS

Data extraction attack

Can we retrieve sensitive data from the training dataset?

Member inference attack

Is this data record part of the training data?

Jailbreak attack

Exploiting the model to gain advantages.

Prompt leakage attack

Leaking the prompt can lead to more successful jailbreak attacks.

MEASURING EFFICIENCY

Time to First Token (TTFT)

Time since request submission until the first token is generated.

Model Load Time

Duration taken to load the model during a cold start.

Token Throughput

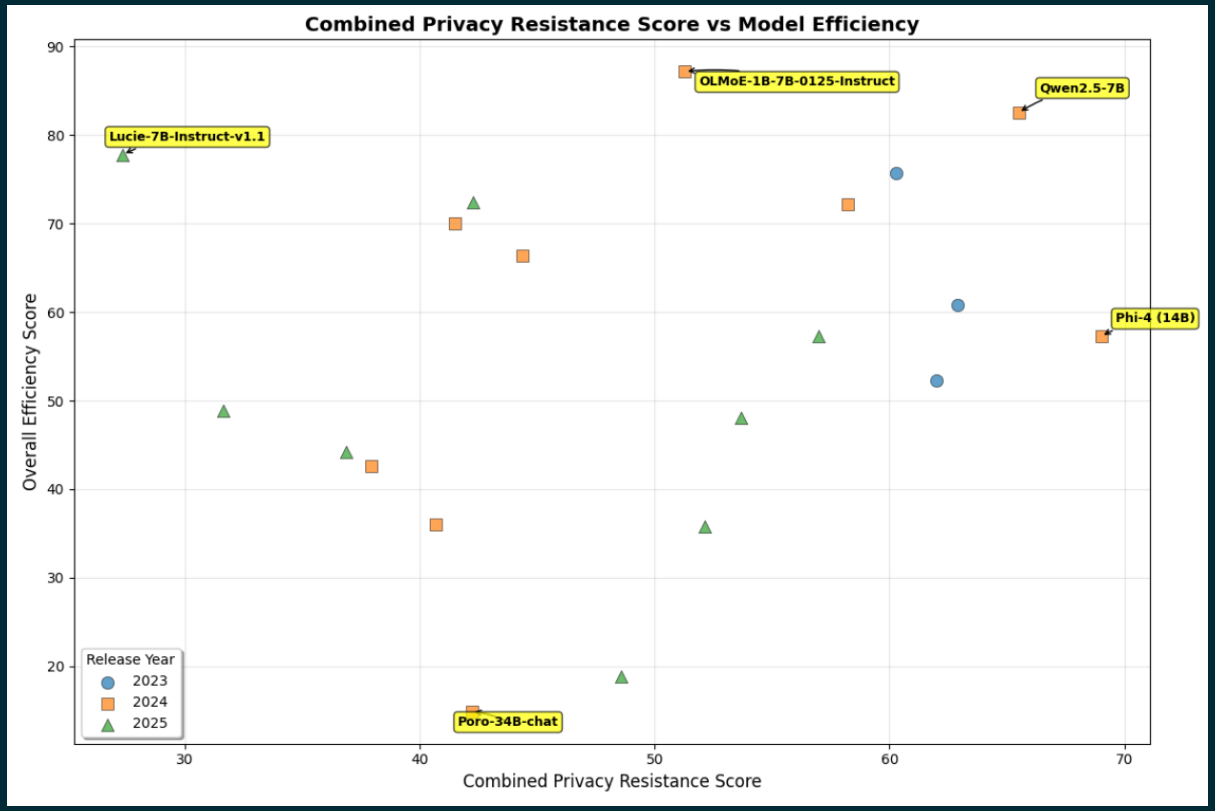
Tokens per second, measured separately for prompt evaluation and generation.

Token Counts

Total amount of tokens processed during prompt and generation phases.

LESSON LEARNED

PRIVACY AND EFFICIENCY ARE NOT MUTUALLY EXCLUSIVE AND CAN BE OPTIMIZED INDEPENDENTLY

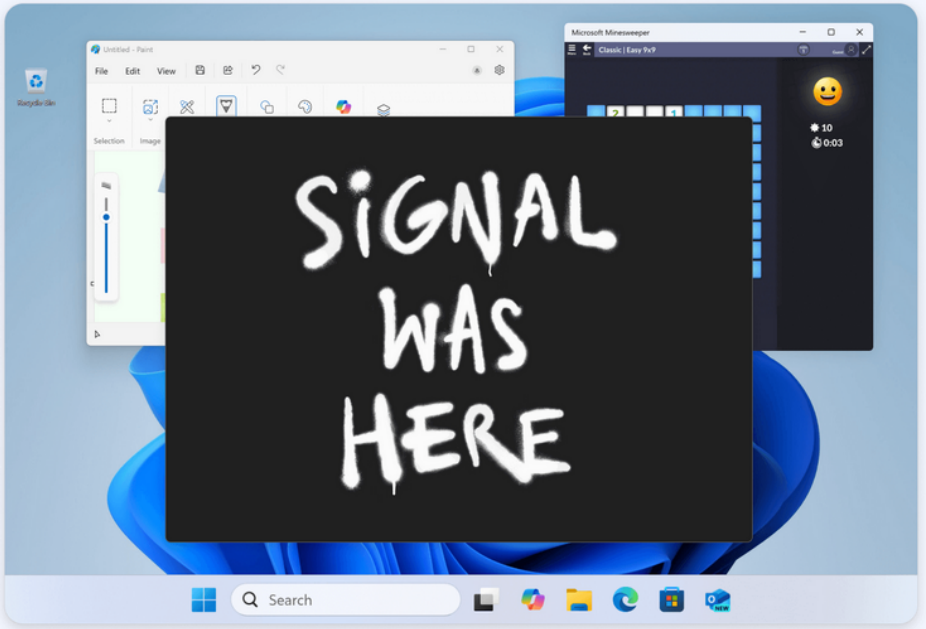


THE DARK SIDE OF AI

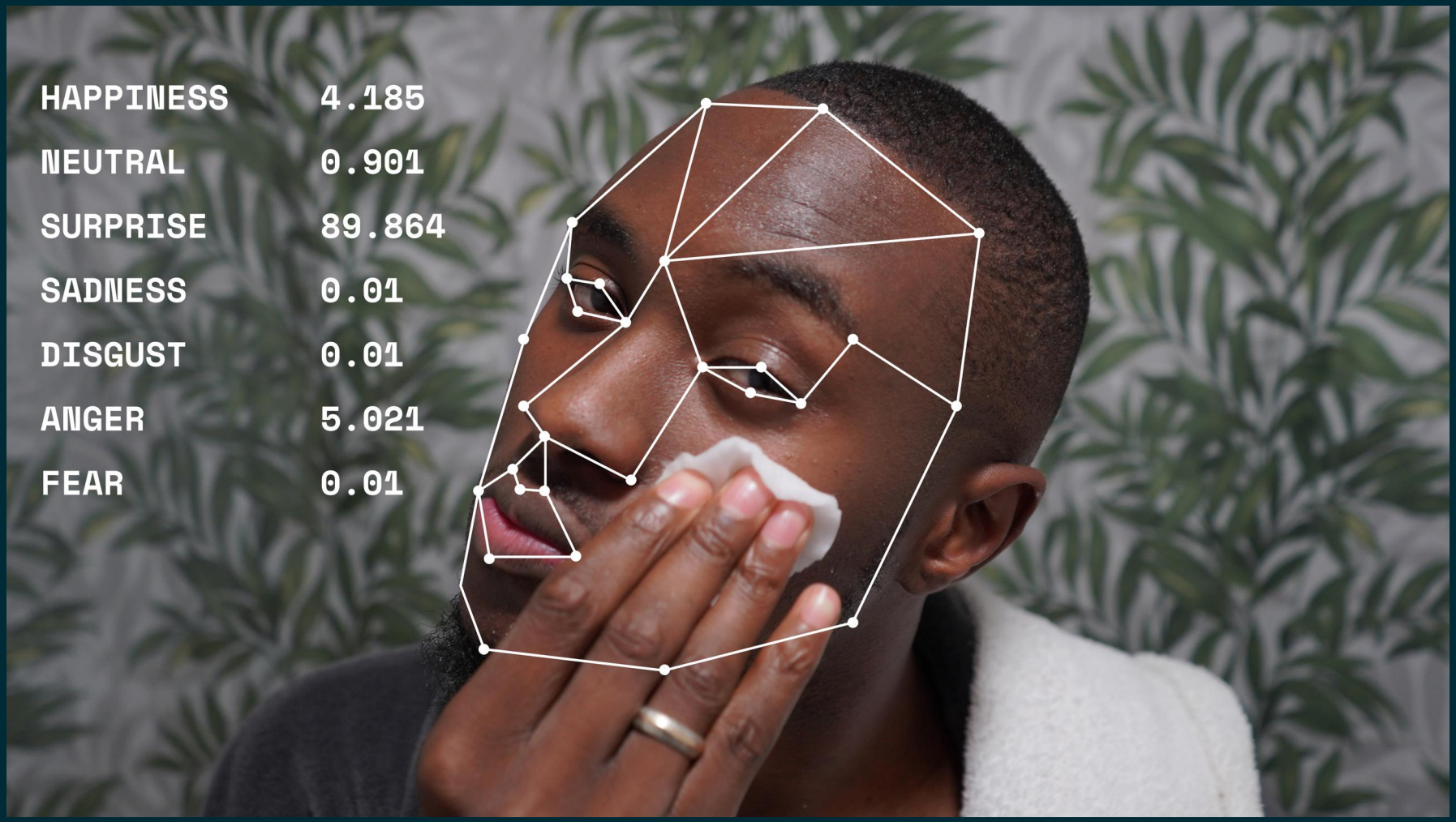
AI AGENTS AT THE OS LEVEL

By Default, Signal Doesn't Recall

jlund on 21 May 2025



ANOTHER EXAMPLE: FACE RECOGNITION



Article

Computer-vision research powers surveillance technology

https://doi.org/10.1038/s41586-025-08972-6

Received: 28 September 2023

Accepted: 3 April 2025

Published online: 25 June 2025

Open access

Check for updates

Pratyusha Ria Kalluri[1]✉, William Agnew[2]✉, Myra Cheng[1], Kentrell Owens[3], Luca Soldaini[4] & Abeba Birhane[5]✉

An increasing number of scholars, policymakers and grassroots communities argue that artificial intelligence (AI) research—and computer-vision research in particular—has become the primary source for developing and powering mass surveillance[1–7]. Yet, the pathways from computer vision to surveillance continue to be contentious. Here we present an empirical account of the nature and extent of the surveillance AI pipeline, showing extensive evidence of the close relationship between the field of computer vision and surveillance. Through an analysis of computer-vision research papers and citing patents, we found that most of these documents enable the targeting of human bodies and body parts. Comparing the 1990s to the 2010s, we observed a fivefold increase in the number of these computer-vision papers linked to downstream surveillance-enabling patents. Additionally, our findings challenge the notion that only a few rogue entities enable surveillance. Rather, we found that the normalization of targeting humans permeates the field. This normalization is especially striking given patterns of obfuscation. We reveal obfuscating language that allows documents to avoid direct mention of targeting humans, for example, by normalizing the referring to of humans as 'objects' to be studied without special consideration. Our results indicate the extensive ties between computer-vision research and surveillance.

TAKEAWAY

FOR SOME CRITICAL APPLICATIONS, WE SHOULD NOT USE AI EVEN IF IT LOOKS USEFUL ON THE SURFACE: IT IS JUST NOT WORTH IT

WRAPPING UP

- When carefully integrated, AI definitely has its uses to enhance privacy
- AI, like any other technology, can be designed in a privacy-friendly way
- Although, for some uses, it can simply lead to a dystopian future and must be avoided

> *"Technology is neither good nor bad; nor is it neutral." Melvin Kranzberg*

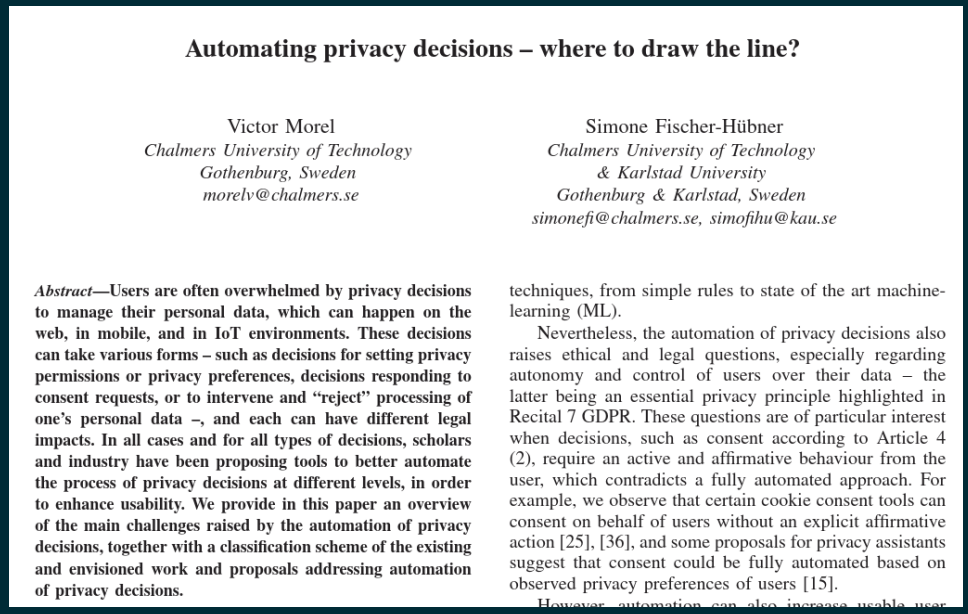
FINAL LESSONS

THINK CAREFULLY ABOUT HOW TO INCLUDE AI IN A SYSTEM

Use privacy-by-design principles.
Leverage Privacy-Enhancing Technologies.
But also question **whether** AI is genuinely required.

APPLICATIONS WHICH REQUIRE CERTAINTY DO NOT BEFIT LLMS

Stochastic output does not go hand in hand with legal certainty.
And sometimes it is not even a question of stochasticity.



Automating privacy decisions – where to draw the line?

Victor Morel
Chalmers University of Technology
Gothenburg, Sweden
morelv@chalmers.se

Simone Fischer-Hübner
Chalmers University of Technology
& Karlstad University
Gothenburg & Karlstad, Sweden
simonefi@chalmers.se, simofihu@kau.se

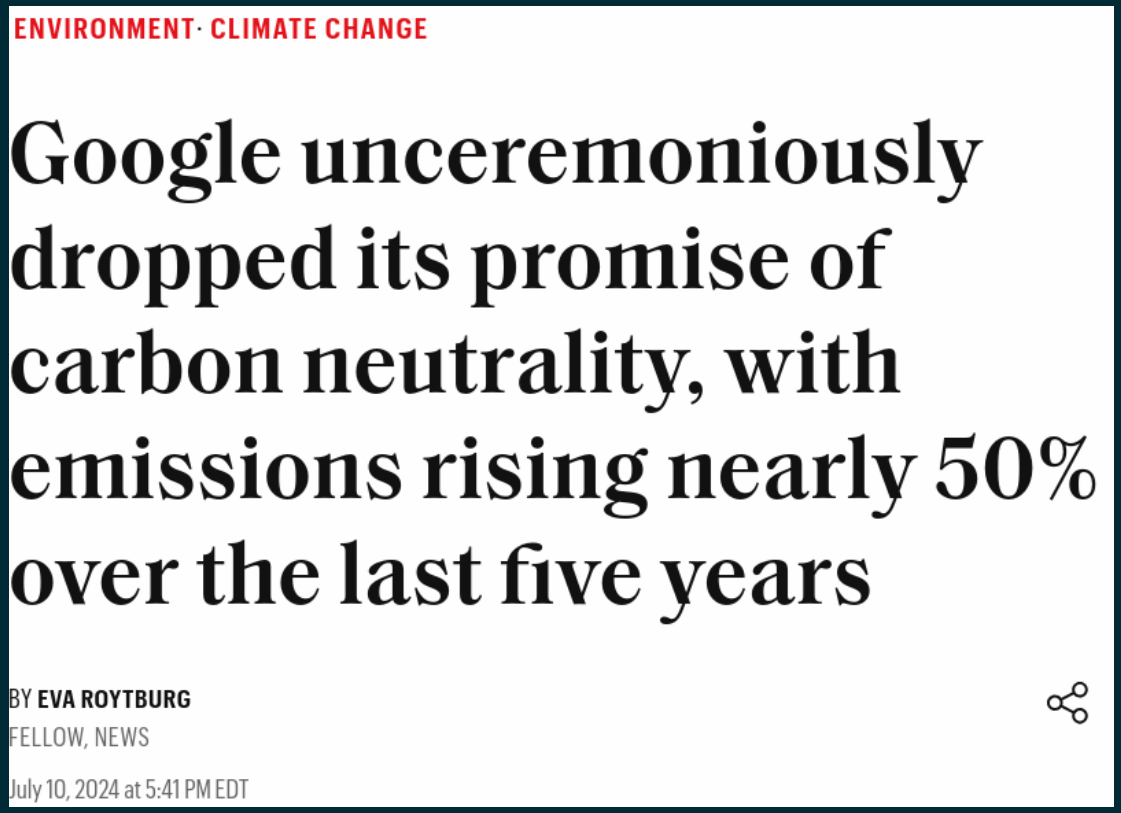
***Abstract*—Users are often overwhelmed by privacy decisions to manage their personal data, which can happen on the web, in mobile, and in IoT environments. These decisions can take various forms – such as decisions for setting privacy permissions or privacy preferences, decisions responding to consent requests, or to intervene and "reject" processing of one's personal data –, and each can have different legal impacts. In all cases and for all types of decisions, scholars and industry have been proposing tools to better automate the process of privacy decisions at different levels, in order to enhance usability. We provide in this paper an overview of the main challenges raised by the automation of privacy decisions, together with a classification scheme of the existing and envisioned work and proposals addressing automation of privacy decisions.**

techniques, from simple rules to state of the art machine-learning (ML).

Nevertheless, the automation of privacy decisions also raises ethical and legal questions, especially regarding autonomy and control of users over their data – the latter being an essential privacy principle highlighted in Recital 7 GDPR. These questions are of particular interest when decisions, such as consent according to Article 4 (2), require an active and affirmative behaviour from the user, which contradicts a fully automated approach. For example, we observe that certain cookie consent tools can consent on behalf of users without an explicit affirmative action [25], [36], and some proposals for privacy assistants suggest that consent could be fully automated based on observed privacy preferences of users [15].

However, automation can also increase usable user

LLMS ALSO HAVE A SIGNIFICANT SOCIAL AND ECOLOGICAL COST



ENVIRONMENT · CLIMATE CHANGE

Google unceremoniously dropped its promise of carbon neutrality, with emissions rising nearly 50% over the last five years

BY **EVA ROYTBURG**
FELLOW, NEWS

July 10, 2024 at 5:41 PM EDT

SOME OF THE CONSIDERATIONS RAISED HERE ALSO APPLY TO SECURITY

- The automation brought by AI can enhance security **although**
- We should focus our efforts to make AI systems secure instead of falling for the hype **and**
- Some critical applications must be avoided if we cannot bring formal proofs

THANKS FOR YOUR ATTENTION

Credits: https://betterimagesofai.org/images