Report from Dagstuhl Seminar 25112

PETs and AI: Privacy Washing and the Need for a PETs Evaluation Framework

Emiliano De Cristofaro[*1], **Kris Shrishak**[*2], **Thorsten Strufe**[*3], **Carmela Troncoso**[*4], **and Felix Morsbach**[†5]

1 **University of California – Riverside, US.** emilianodc@cs.ucr.edu
2 **Irish Council for Civil Liberties – Dublin, IE.** kris.shrishak@iccl.ie
3 **KIT – Karlsruher Institut für Technologie, DE.** thorsten.strufe@kit.edu
4 **MPI-SP – Bochum, DE.** carmela.troncoso@mpi-sp.org
5 **KIT – Karlsruher Institut für Technologie, DE.** felix.morsbach@kit.edu

Abstract

As public awareness of data collection practices and regulatory frameworks grows, privacy-enhancing technologies (PETs) have emerged as a promising approach to reconciling data utility with individual privacy rights. PETs underpin privacy-preserving machine learning (PPML), integrating tools like differential privacy, homomorphic encryption, and secure multiparty computation to safeguard data throughout the AI lifecycle. However, despite significant technical progress, PETs face critical policy and governance challenges. Recent works have raised concerns about efficacy and deployment of PETs, observing that fundamental rights of people are continually being harmed, including, paradoxically, privacy. PETs have been used in surveillance applications and as a privacy washing tool. Current approaches often fail to address broader harms beyond data protection, highlighting the need for a more comprehensive privacy evaluation framework. This Dagstuhl Seminar brought together scholars in computer science and law, along with policymakers, regulators, and industry leaders, to discuss privacy washing and the challenges of detecting privacy washing through PETs and explored pathways toward a framework to address these challenges.

Seminar March 9–14, 2025 – https://www.dagstuhl.de/25112

2012 ACM Subject Classification Security and privacy → Privacy protections; Security and privacy → Privacy-preserving protocols; Security and privacy → Social aspects of security and privacy

Keywords and phrases Privacy Enhancing Technologies (PET), Privacy Evaluation, Privacy Harm, Privacy Threats, Privacy Washing

Digital Object Identifier 10.4230/DagRep.15.3.77

1 Executive Summary

Emiliano De Cristofaro (University of California – Riverside, US)
Kris Shrishak (Irish Council for Civil Liberties – Dublin, IE)
Thorsten Strufe (KIT – Karlsruher Institut für Technologie, DE)
Carmela Troncoso (MPI-SP – Bochum, DE)

Privacy is a fundamental human right. Article 12 of the Universal Declaration of Human Rights (UDHR) states that everyone has the right to potection from interference with their privacy. One part of protecting people's privacy is data protection. Laws such as the EU's

[*] Editor / Organizer
[†] Editorial Assistant / Collector

PETs and AI: Privacy Washing and the Need for a PETs Evaluation Framework, *Dagstuhl Reports*, Vol. 15, Issue 3, pp. 77–93
Editors: Emiliano De Cristofaro, Kris Shrishak, Thorsten Strufe, Carmela Troncoso, and Felix Morsbach
DAGSTUHL REPORTS Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

General Data Protection Regulation (GDPR) have been drafted to protect personal data, which can be exploited to interfere with people's private life. Numerous countries around the world have adopted laws similar to the GDPR. These laws along with an increased awareness of personal data collection have contributed to the appeal of technological solutions known broadly as privacy enhancing technologies (PETs).

The premise of PETs is that these techniques allow data processing while protecting the underlying data from being revealed unnecessarily. PETs make it possible to analyse data from multiple sources without having to see the data. There are two major kinds of PETs: one that offers input privacy and another that offers output privacy. Input privacy allows different people to pour-in their individual data to combine and generate an insight, while no one learns anyone else's individual data. For example, a group of friends can learn who earns the highest without revealing their individual salary to each other. Techniques such as homomorphic encryption and secure multiparty computation (SMPC) fall into this category. These powerful techniques allow two or more entities to compute an agreed upon function on encrypted data. They are useful when the participating entities do not trust each other with their private inputs, but see mutual benefit in the output of the function. Output privacy allows for the release of aggregate data and statistical information while preventing the identification of individuals. Techniques such as differential privacy fall into this category. In many practical use cases, both input and output privacy is desired, and these techniques are combined.

One such use case is AI and in particular machine learning (ML). A huge volume of data is a key component of some of the machine learning techniques, especially those relying on deep neural networks. Personal data and those with sensitive attributes are also used to develop AI models. However, this has contributed to privacy risks. There are a range of attacks in the literature that aim to extract personal data from trained models. PETs have been proposed as the way to protect the functionality of AI while protecting against these privacy attacks. In fact, an entire research field known as privacy-preserving machine learning (PPML) has been formed. PPML incorporates various PETs techniques at various stages of the machine learning to (a) train over encrypted data (e.g., with homomorphic encryption or SMPC), (b) anonymize training process (e.g., DP-SGD), and (c) protect the outputs using differential privacy.

Despite the abundance of works in the area of PETs, AI, and their intersection, there are many remaining challenges. Addressing these challenges is crucial to understand the drawbacks and to reap the benefits of PETs. A range of research questions in Computer Science (protocol design, privacy guarantees, feasibility, scalability, efficiency, etc.) need to be addressed. There are also questions that are interdisciplinary and require expertise from NGOs, ethicists, policy making, law, and regulators. And these research questions are not merely to satisfy academic curiosity but have practical ramifications. They could affect policy making and the work of regulators.

In this Dagstuhl Seminar, a multidisciplinary group of computer science and legal academics and practitioners from industry, human rights groups, and regulators discussed two challenges:

1. **Privacy washing through PETs**: In the recent years, PETs have been used in surveillance applications as in the case of Apple's proposed (and then retracted) approach to scan images on people's phones when uploading photos to iCloud. They have also been used in applications where the personal data is seemingly protected but the privacy threats faced by people are amplified, for example in targeted advertising. Such applications show that PETs can be used for "privacy washing". At the heart of the issue is that

most works fail to protect against the interference with privacy as laid down in Article 12 of the UDHR. These works are agnostic to the application context or too generic or limited to the cryptographic protocol without considering the privacy threats due to the system where it is embedded. The imbalances and asymmetries of power between the stakeholders, the role of infrastructures and their providers, and the control of the computing infrastructure are not accounted for. Technical measures to protect data are discussed as being equivalent to privacy, when they are not. Privacy violations can take many other forms including economic and discrimination harms. When the goal of the application is to harm privacy, such technical measures to protect data cannot protect the interference with privacy. The threat models in the literature are inadequate, and thus, systems designed under such models continue to cause privacy harms.

2. **Evaluation framework to detect privacy washing**: If PETs are to protect against interference with privacy, as laid down in the UDHR, then we require standard evaluation methods and frameworks that allow us to compare the degree of protection. While the literature is filled with ways to measure PETs, they are hard to compare. Limitations of PETs should be well documented so that privacy washing through PETs is stopped. A lack of an independent evaluation framework allows privacy washing. Addressing this challenge is timely and this seminar took the initial steps towards an evaluation framework.

Seminar Structure

Since the participants came from diverse backgrounds ranging from different topics in computer science to legal and regulatory work, the seminar began with several introductory talks and two panel discussions to bring everyone up to speed. Then, we brainstormed in small groups about all the aspects that could influence whether the deployment of a PET could be considered privacy washing. We subsequently grouped these aspects into four topics: Functionality and Framing, Infrastructure for PETs, Accountability, and Detection of Fake PETs. We split the group into four subgroups to discuss these aspects further and develop criteria by which to evaluate the deployment of a PET leading to a vast catalogue of factors that influence the efficacy of PET deployments. During the plenary meetings after group discussions, the rapporteurs from each group shared the progress made during the group discussions. Finally, we spent the remaining time to merge the results of the four subgroups into a draft for a position paper. The position paper describes what privacy washing is, who is involved in its deployment, who can be affected by it, and the considerations that help to detect privacy washing in deployed systems.

2 Table of Contents

Executive Summary

Overview of Talks

Panel discussions

Working groups

3 Overview of Talks

3.1 Introduction to Differential Privacy and Federated Learning

Aurélien Bellet (INRIA – Montpellier, FR)

In an era of AI-driven applications, balancing data utility with user privacy is more important than ever. This talk provides a high level introduction to two key approaches addressing this challenge: federated learning and differential privacy. Federated learning enables collaborative model training without sharing raw data, while differential privacy provides strong guarantees against individual data leakage. This talk discusses the fundamental ideas behind these techniques, their real-world applications, and some challenges that remain.

3.2 Agency Protection: Organisms & Institutions

Robin Berjon (Princeton, US)

It's always tempting to cut things at what seem like logical joints so as to make thinking about the individual component easier. In a sense, that's what we've done with privacy, focusing primarily on various forms of data processing. But the world is rarely as orthogonal as we model it to be. This brief talk situated privacy in a wider institutional framework and suggests that we may use an institutional grammar to evaluate the role and effectiveness of privacy decisions in a broader context.

3.3 Purpose formulations as a weak link in data protection

Asia Biega (MPI-SP – Bochum, DE)

Purpose limitation is one of the requirements under the GDPR. User data has to be processed for specific, explicit, and legitimate purposes. Purpose formulations specify and describe these purposes. In this talk, I presented four examples that, over time, convinced me that these formulations are a weak link in data protection: and thus become a tool for privacy washing.

3.4 A regulator’s perspective on data protection and privacy

Paul Comerford (Information Commissioner’s Office – Wilmslow, GB)

Paul Comerford (Principal Technology Adviser at the ICO) discussed the role of the technology and innovation directorate at the ICO. The talk focused on the ICOs work on PETs across multiple domains and its upcoming guidance on anonymisation and pseudonymisation. He also discussed our recent work on AI and PETs.

3.5 Anonymisation: Introduction and Perspectives

Ana-Maria Cretu (EPFL – Lausanne, CH)

Main reference Andrea Gadotti, Luc Rocher, Florimond Houssiau, Ana-Maria Creţu, Yves-Alexandre de Montjoye: “Anonymization: The imperfect science of using data while preserving privacy”, Science Advances, Vol. 10(29), p. eadn7053, 2024.
URL https://doi.org/10.1126/sciadv.adn7053

Anonymisation is the main legal paradigm for sharing data while protecting people’s right to privacy. In spite of decades of research, robust anonymisation (“de-identifying”) of individual-level datasets remains an elusive goal. Numerous re-identification attacks have indeed shown how adversaries can use auxiliary information about individuals to single them out in supposedly anonymous datasets. One solution to the data sharing problem is aggregation, whereby data owners share with third parties the results of a computation across all records, while retaining control over the individual-level data. Aggregation solutions include summary statistics, interactive queries over the data, synthetic data, and machine learning. But aggregation does not, on its own, protect privacy, and evaluating the privacy of these solutions is far from trivial. This talk described the two main approaches for this: (1) designing and evaluating privacy attacks and (2) formal methods based on differential privacy, with their advantages and their challenges, together with my perspective on the field.

3.6 Attacks on privacy-preserving systems

Yves-Alexandre de Montjoye (Imperial College London, GB)

Companies and governments are increasingly relying on privacy-preserving techniques to collect and process sensitive data. In this talk, I will discuss our efforts to red team deployed systems and argue that red teaming is essential to protect privacy in practice. I will first shortly describe how traditional de-identification techniques fail in today’s world. I will then show how implementation choices and trade-offs have enabled attacks against real-world systems, from query-based systems to differential privacy mechanisms and synthetic data. I will conclude by discussing how this applies to modern AI systems.

3.7 Understanding and addressing fairwashing in machine learning

Sébastien Gambs (UQAM – Montreal, CA)

Fairwashing refers to the risk that an unfair black-box model can be explained by a fairer model through post-hoc explanation manipulation. In this talk, I will first discuss how fairwashing attacks can transfer across black-box models, meaning that other black-box models can perform fairwashing without explicitly using their predictions. This generalization and transferability of fairwashing attacks imply that their detection will be difficult in practice. Finally, I will nonetheless review some possible avenues of research on how to limit the potential for fairwashing.

3.8 The PET Paradox – The case of Amazon Sidewalk

Seda F. Gürses (TU Delft, NL)

Main reference Thijmen van Gend, Donald Jay Bertulfo, Seda F. Gürses: "The PET Paradox: How Amazon Instrumentalises PETs in Sidewalk to Entrench Its Infrastructural Power", CoRR, Vol. abs/2412.09994, 2024.
URL https://doi.org/10.48550/ARXIV.2412.09994

Recent applications of Privacy Enhancing Technologies (PETs) reveal a paradox. PETs aim to alleviate power asymmetries, but can actually entrench the infrastructural power of companies implementing them vis-à-vis other public and private organizations. We investigate whether and how this contradiction manifests with an empirical study of Amazon's cloud connectivity service called Sidewalk. In 2021, Amazon remotely updated Echo and Ring devices in consumers' homes, to transform them into Sidewalk "gateways". Compatible Internet of Things (IoT) devices, called "endpoints", can connect to an associated "Application Server" in Amazon Web Services (AWS) through these gateways. We find that Sidewalk is not just a connectivity service, but an extension of Amazon's cloud infrastructure as a software production environment for IoT manufacturers. PETs play a prominent role in this pursuit: we observe a two-faceted PET paradox. First, suppressing some information flows allows Amazon to promise narrow privacy guarantees to owners of Echo and Ring devices when "flipping" them into gateways. Once flipped, these gateways constitute a crowdsourced connectivity infrastructure that covers 90% of the US population and expands their AWS offerings. We show how novel information flows, enabled by Sidewalk connectivity, raise greater surveillance and competition concerns. Second, Amazon governs the implementation of these PETs, requiring manufacturers to adjust their device hardware, operating system and software; cloud use; factory lines; and organizational processes. Together, these changes turn manufacturers' endpoints into accessories of Amazon's computational infrastructure; further entrenching Amazon's infrastructural power. We discuss similarities and differences between previous strategic uses of PETs by Google and Apple to expand their infrastructural offerings to third parties. Accordingly, we argue that power analyses undergirding PET designs should go beyond analyzing information flows. We propose future steps for policy and tech research.

3.9 PETs Intro: Multiparty Computation and Homomorphic Encryption

Bailey Kacsmar (University of Alberta – Edmonton, CA)

In this session we provided an overview on what multiparty computation (MPC) is and how we can think about its variants. We similarly discussed homomorphic encryption (HE). The goal with this session was to establish the breadth of the areas and provide attendees with a common language to think about the way privacy enhancing technologies (PETs) that employ MPC and HE can vary; allowing us to better evaluate the implications of these technologies for privacy and artificial intelligence. We concluded with an overview of some of what is currently possible, in terms of applications, that employ MPC and HE.

3.10 What I believe privacy engineering is and some missing pieces

Carmela Troncoso (MPI-SP – Bochum, DE)

In this talk we revisited previous definitions of privacy engineering, showing that data or trust minimization do not necessarily minimize harms. We then argue that purpose minimization is the design goal that helps in this respect. Purpose-oriented thinking additionally has a benefit that it enables to identify fundamental purposes and harms that derive from the goal of the system and have to be assumed as a risk should the system be deployed. We then discussed some missing definitions that would allow to capture harms associated to function creep.

3.11 You Still See Me

Rui-Jie Yew (Brown University – Providence, US)

Main reference Rui-Jie Yew, Lucy Qin, Suresh Venkatasubramanian: "You Still See Me: How Data Protection Supports the Architecture of ML Surveillance", CoRR, Vol. abs/2402.06609, 2024.
URL https://doi.org/10.48550/ARXIV.2402.06609

Data forms the backbone of artificial intelligence (AI). Privacy and data protection laws thus have strong bearing on AI systems. Shielded by the rhetoric of compliance with data protection and privacy regulations, privacy-preserving techniques have enabled the extraction of more and new forms of data. In this talk, I illustrate how the application of privacy-preserving techniques in the development of AI systems–from private set intersection as part of dataset curation to homomorphic encryption and federated learning as part of model computation–can further support surveillance infrastructure under the guise of regulatory permissibility. Finally, I propose technology and policy strategies to evaluate privacy-preserving techniques in light of the protections they actually confer. I conclude by highlighting the role that technologists can play in devising policies that combat surveillance AI technologies.

4 Panel discussions

4.1 AI models, data protection and privacy washing

Aurélien Bellet (INRIA – Montpellier, FR), Asia Biega (MPI-SP – Bochum, DE), Paul Comerford (Information Commissioner's Office – Wilmslow, GB), Yves-Alexandre de Montjoye (Imperial College London, GB), Carmela Troncoso (MPI-SP – Bochum, DE), and Rui-Jie Yew (Brown University – Providence, US)

The panel explored how privacy-enhancing technologies (PETs) and regulatory tools are increasingly used for privacy washing – creating a surface-level appearance of compliance while sidestepping real accountability. Sandboxes and red teaming were called out as processes that can be used for legitimizing privacy-invasive systems without addressing underlying risks. Technologies like differential privacy, synthetic data generation and federated learning were highlighted as particularly vulnerable to misuse, especially when their implementation details are obscured or when their guarantees are undermined through practices like budget resetting or general misconfigurations. A key point raised was that evaluations should prioritize the actual impact on individuals and society, not just technical compliance or claimed adherence to norms.

The conversation also focused on the role and limits of transparency. While transparency was broadly supported as essential for accountability, it was acknowledged that legal barriers like trade secrets and competition law often prevent meaningful oversight. There was an agreement that transparency should go beyond abstract metrics and provide explanations that are intelligible to non-experts. At the same time, concerns were raised that transparency alone can also be co-opted as another form of privacy washing if not paired with enforcement and verification. The discussion underscored the need for enforceable standards, empowered regulators, and a shift away from over-optimizing technical frameworks toward addressing broader systemic and structural issues in privacy governance.

4.2 User Perspective of PETs

Bailey Kacsmar (University of Alberta – Edmonton, CA), Robin Berjon (Princeton, US), Emiliano De Cristofaro (University of California – Riverside, US), Lucy Qin (Georgetown University – Washington, DC, US), and Carmela Troncoso (MPI-SP – Bochum, DE)

The panel discussed the gap between how privacy-enhancing technologies (PETs) are developed and how real users understand, need, or experience them. There was a recurring argument that users often lack the language, awareness, or mental models to demand privacy – much like people once lacked the concept of clean tap water – yet that doesn't mean privacy isn't essential. PETs should be designed to be invisible and default, not something users must consciously engage with. Communication breakdowns between researchers, usability experts, and end-users were identified as major barriers. There was also a push to broaden the definition of "users" to include software engineers and institutional actors, since engineers are often key decision-makers and operate much closer to the tools in practice. The lack of actionable usability research and the assumption of a clearly defined privacy "problem" were cited as weaknesses in the current ecosystem.

Beyond end-users, the conversation highlighted the role of high-risk populations, NGOs, policymakers, and businesses in the PETs landscape. Messaging must be tailored – migrants, for instance, face urgent harms that don't always register as "privacy" risks. While PETs can support collective systems like digital commons, structural components and effective messaging are missing. Some companies adopt PETs reactively (e.g. post-GDPR), while others see them as a branding opportunity – but distinguishing meaningful implementations from superficial ones remains difficult. There's also underused potential in inter-organizational PET deployments and in rethinking how to engage businesses without falling into technical "impossibility" traps. A key takeaway: users shouldn't bear the burden of privacy, and communicating harm – especially to those at risk – must be better informed, more targeted, and more pragmatic.

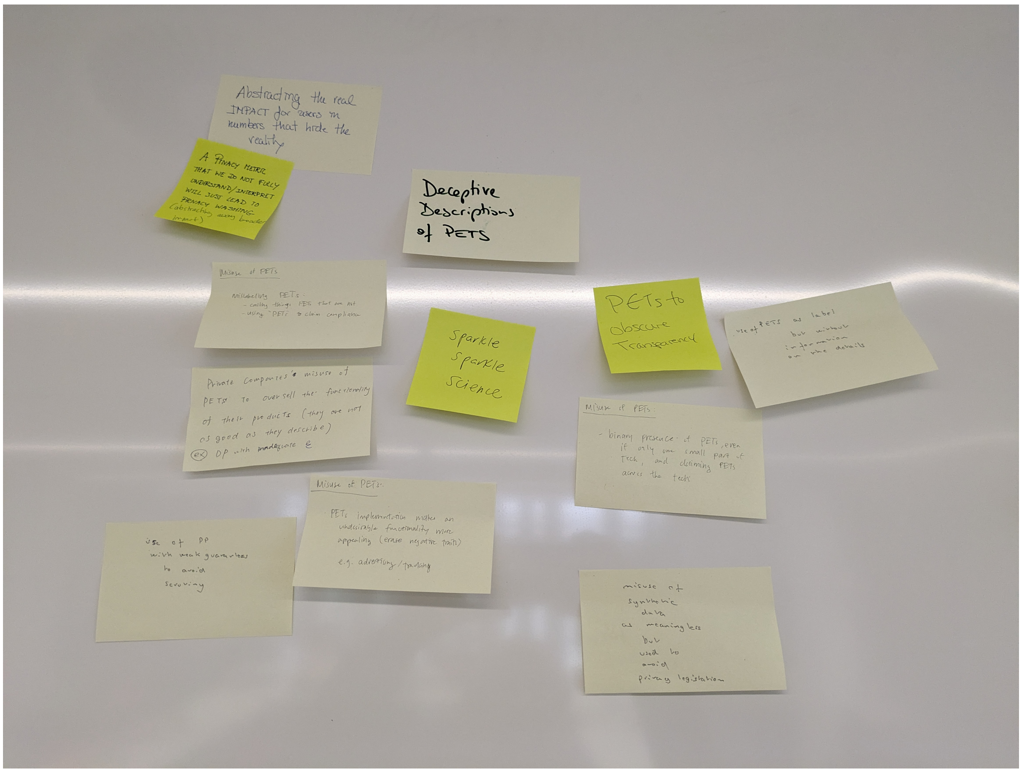
5 Working groups

5.1 Detecting Fake PETs

Frederik Armknecht (Universität Mannheim, DE), Aurélien Bellet (INRIA – Montpellier, FR), Ana-Maria Cretu (EPFL – Lausanne, CH), Yves-Alexandre de Montjoye (Imperial College London, GB), Georgi Ganev (University College London, GB), Patricia Guerra-Balboa (KIT – Karlsruher Institut für Technologie, DE), Felix Morsbach (KIT – Karlsruher Institut für Technologie, DE), and Thorsten Strufe (KIT – Karlsruher Institut für Technologie, DE)

The group focused on defining a structured approach to detect fake PETs – privacy-enhancing technologies that mislead through exaggerated claims, poor implementation, or misconfiguration. A central proposal was to create a standardized transparency tool, akin to model cards or data sheets, tentatively referred to as a "privacy card." This would contain minimal but essential information to assess whether a system is making valid privacy claims. The discussion outlined four key failure categories in PETs: mismatch (between claims and actual threat mitigation), overestimation (inflated protection claims), wrong implementation, and wrong configuration. A foundational requirement is that any privacy claim must specify the threat model, the PET used, and the degree of mitigation. Vague claims without a clear adversarial context were identified as a red flag and sufficient grounds for labeling the system a fake PET.

Each failure mode was elaborated with practical evaluation steps. For mismatch and overestimation, the group emphasized decomposing systems into discrete threat-mitigation claims and validating them with formal or empirical evidence. Identifying implementation failures requires either open-source access or a reproducible protocol, with particular scrutiny on subtleties like side channels, flawed randomness, or deviations from trusted primitives. Configuration errors, such as excessive ϵ values in differential privacy or undersized encryption keys, must be contextualized within both the system's technical parameters and its deployment environment. The group stressed that privacy guarantees are only meaningful when technical claims are precise, verifiable, and aligned with real-world adversary models – making transparent, auditable documentation a practical necessity to prevent privacy washing.

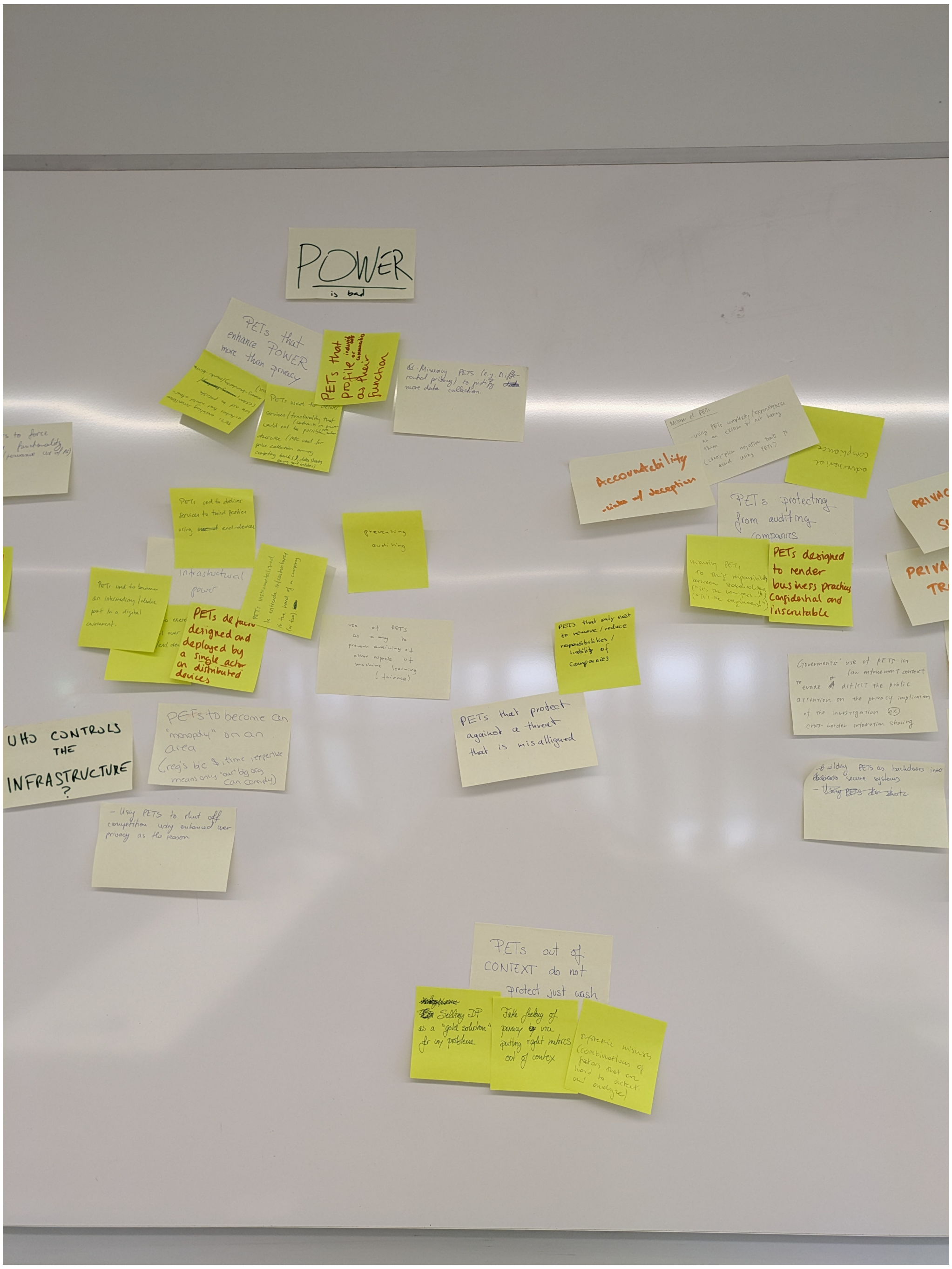

Abstracting the real IMPACT for users in numbers that hide the reality
A PRIVACY METRIC THAT WE DO NOT FULLY UNDERSTAND/INTERPRET WILL JUST LEAD TO PRIVACY WASHING (abstracting away broader impact)
Deceptive Descriptions of PETS
Misuse of PETs
Mislabeling PETs:
- calling things PETs that are not
- using "PETs" to claim compliance
Sparkle Sparkle Science
PETs to obscure Transparency
use of PETS as label but without information on the details
Private Companies' misuse of PETS to oversell the functionality of their products (they are not as good as they describe)
ex DP with inadequate ε
Misuse of PETs:
- binary presence of PETs, even if only one small part of Tech, and claiming PETs across the tech
Misuse of PETs:
PETs implementation makes an undesirable functionality more appealing (erase negative traits)
e.g. advertising/tracking
use of PP with weak guarantees to avoid scrutiny
misuse of synthetic data as meaningless but used to avoid privacy legislation

5.2 Infrastructure & PETs

Robin Berjon (Princeton, US), Seda F. Gürses (TU Delft, NL), Lucy Qin (Georgetown University – Washington, DC, US), Michael Veale (University College London, GB), and Rui-Jie Yew (Brown University – Providence, US)

The discussion highlighted that evaluating Privacy-Enhancing Technologies (PETs) cannot be isolated from the computational infrastructure they rely on, which often embodies extractive or privacy-compromising characteristics. A key challenge identified is the "stack problem": PETs depend on underlying infrastructures that may themselves lack privacy protections, making truly independent PETs difficult to build and sustain without relying on PET-compatible infrastructure, governance, and funding. This dynamic concentrates power among well-resourced entities capable of controlling infrastructure, raising concerns about exclusionary effects on who can develop or maintain PETs based on existing economic and political incentives.

The group further emphasized that the production environment and infrastructure shape PET design, deployment, and sustainability, often introducing trust relationships and operational vulnerabilities. Coordination among infrastructure providers, deployers, and users creates new power relations, sometimes consolidating rather than distributing it. Ultimately, privacy-washing occurs when infrastructural dependencies and power asymmetries are overlooked, leading to overstated claims about PET's protections while entrenching systemic privacy risks. Effective evaluation frameworks must therefore assess PETs in a full-stack context, including the socio-technical and governance layers that support or constrain them.

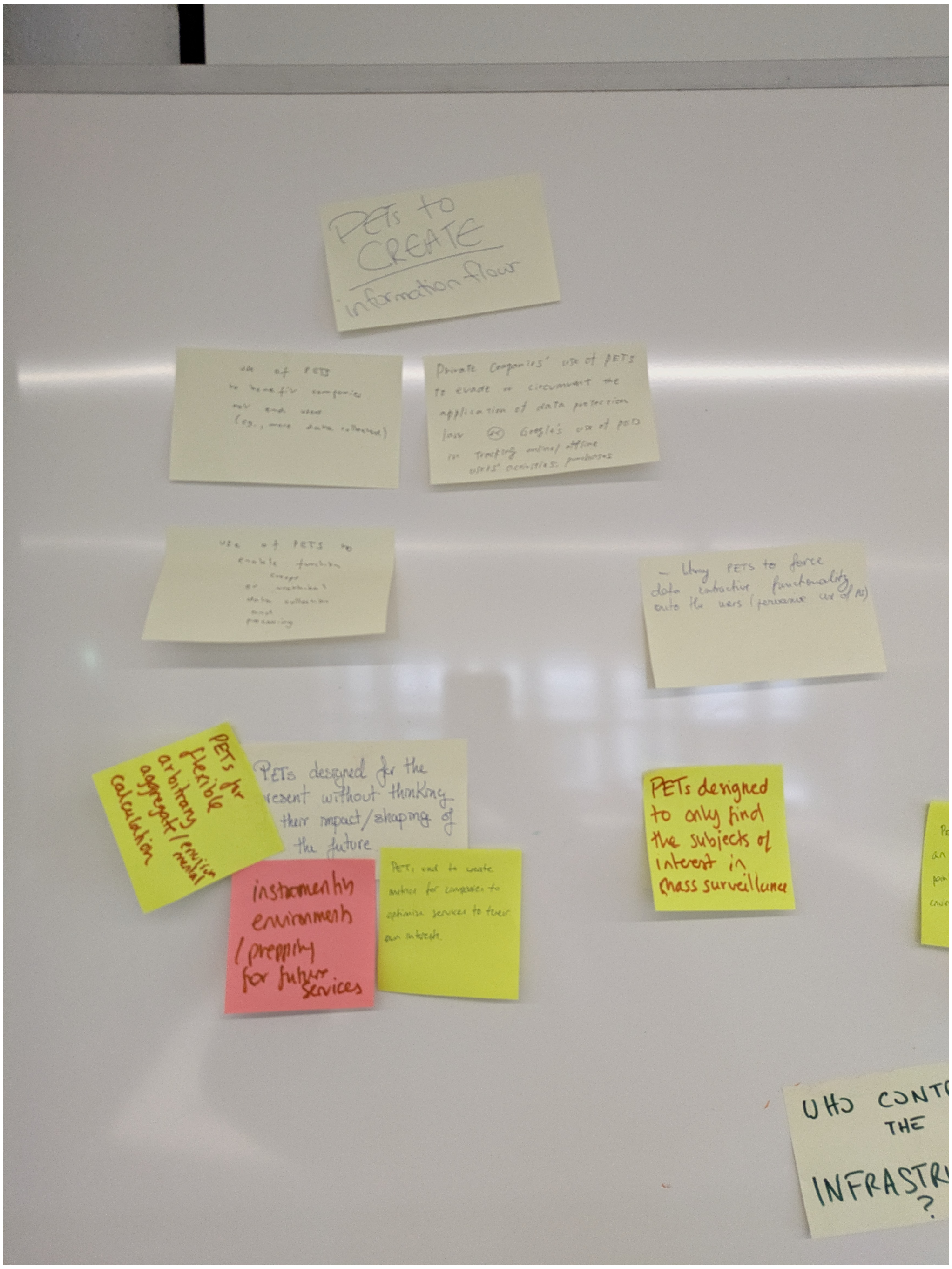

POWER
is bad
PETs that enhance POWER more than privacy
PETs that profile individuals or communities as their function
Misusing PETS (e.g. Differential privacy) to justify more data collection.
PETs used to deliver services / functionality that would not be possible otherwise (constraints on power are broken) (MPC used for price collusion among competing banks, data sharing among govt entities)
PETs to force functionality (pervasive use of AI)
PETs used to deliver services to third parties using end-devices
preventing auditing
infrastructural power
PETs used to become an intermediary / choke point in a digital environment.
PETs de facto designed and deployed by a single actor on distributed devices
PETs instrumentalized to entrench infrastructure in the hand of a company (or two).
use of PETS as a way to prevent auditing of other aspects of machine learning (fairness)
WHO CONTROLS THE INFRASTRUCTURE ?
PETs to become an "monopoly" on an area (reg's b/c $, time, expertise means only "our" big org can comply)
- Using PETS to shut off competition using enhanced user privacy as the reason
PETs that protect against a threat that is misalligned
PETs that only exist to remove / reduce responsibilities / liability of companies
Accountability - sinks of deception
Misuse of PETs
- using PETs complexity / expensiveness as an excuse for not having them
(cherry-pick negative traits to avoid using PETs)
adversarial compliance
PETs protecting from auditing companies
misusing PET, to shift responsibility between stakeholders ("it's the lawyers...!") ("it's the engineers!")
PETs designed to render business practices confidential and inscrutable
PRIVAC
PRIVA
TR
Goverments' use of PETS in law enforcement context To evade or deflect the public attention on the privacy implication of the investigation ex cross-border information sharing
- Building PETS as backdoors into secure systems
- ~~Using PETS to~~
PETs out of CONTEXT do not protect just wash
Selling DP as a "gold solution" for any problems
Fake feeling of privacy via putting right metrics out of context
systemic misuses (combinations of factors that are hard to detect and analyze)

5.3 Functionality and Framing

Asia Biega (MPI-SP – Bochum, DE), Johanna Gunawan (Maastricht University, NL), Hinako Sugiyama (University of California – Irvine, US), Vanessa Teague (Australian National University – Acton, AU), and Carmela Troncoso (MPI-SP – Bochum, DE)

The group explored how privacy-enhancing technologies (PETs) can be co-opted to obscure harm rather than mitigate it, calling for a taxonomy of privacy-washing methods to support clearer evaluation, from the system's purpose, to its implementation, communication, and resulting consequences. Three key forms of privacy washing were identified: first, when PETs are layered over systems whose underlying purpose is harmful or objectionable, PETs cannot fix this inherent harm; second, when the system's implementation is harmful, even if its purpose is legitimate, and PETs are used to mask this; and third, when misleading communication about PETs causes harm – such as falsely marketing systems as end-to-end encrypted. In all three cases, PETs risk being used as decorative compliance tools, deflecting attention from structural issues or enabling more sophisticated forms of manipulation, profiling, or opacity.

The group proposed analyzing PETs through the full lifecycle of a system: from purpose, to technical implementation, to communication, and finally to consequences. A system-wide framing was emphasized – assessing not just whether a PET works, but whether it genuinely addresses the privacy risks tied to the system's function and context. PETs that enable or justify harmful practices and information flows were flagged as particularly concerning. The group cautioned against starting privacy assessments too late in the process (e.g., at the DPIA stage), noting that foundational design choices may already predetermine harm. A meaningful framework, they argued, must account for both intentional and structural misuses of PETs, especially as these tools are increasingly used to manage – not eliminate – power imbalances and information asymmetries.

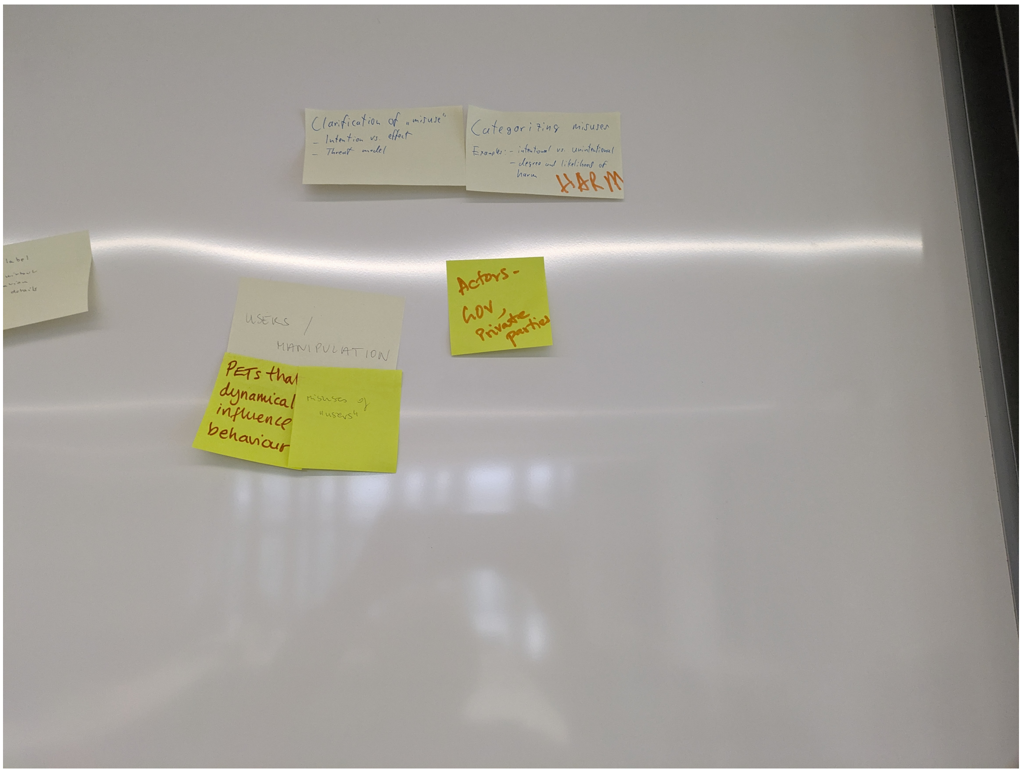

PETs to CREATE
information flow
use of PETS to benefit companies not end user (e.g., more data collected)
Private Companies' use of PETS to evade or circumvent the application of data protection law ex Google's use of PETS in tracking online/offline users' activities, purchases
use of PETS to enable function except or unethical data collection and processing
– Using PETS to force data extractive functionality onto the users (pervasive use of AI)
PETs for flexible arbitrary aggregate/environmental calculation
PETs designed for the present without thinking their impact/shaping of the future
instrumenting environments / prepping for future services
PETs used to create metrics for companies to optimize services to their own interests.
PETs designed to only find the subjects of interest in mass surveillance
WHO CONT
THE
INFRASTR
?

5.4 Accountability

Bailey Kacsmar (University of Alberta – Edmonton, CA), Paul Comerford (Information Commissioner's Office – Wilmslow, GB), Sébastien Gambs (UQAM – Montreal, CA), and Kris Shrishak (Irish Council for Civil Liberties – Dublin, IE)

The group discussed how PETs can be misused to block accountability, particularly by preventing audits, obscuring system behavior, or undermining data access rights. Rather than supporting transparency, some PETs are designed – or framed – as privacy solutions while enabling organizations to evade scrutiny. For example, this includes using PETs to justify blocking data subject's access rights under the GDPR, offloading privacy responsibilities to third-party service providers, or deploying unverifiable systems that require trust without oversight. The group emphasized the importance of designing PETs with auditability and verifiability in mind to counter these failures and ensure they contribute to, rather than hinder, accountability.

In the context of AI, similar dynamics emerge. Techniques like federated learning – while often cited as privacy-preserving – can be used to obscure data processing practices and resist evaluation due to their complexity. Participants noted that organizations may deploy PETs to legitimize questionable practices or bypass legal requirements, especially in the private sector where business incentives dominate. Overall, the discussion called for evaluation frameworks that address how PETs are used in practice – focusing not just on their technical properties, but also on their role in enabling or obstructing rights, oversight, accountability, and public trust.


Clarification of „misuse"
– Intention vs. effect
– Threat model
Categorizing misuses
Examples: – intentional vs. unintentional
– degree and likelihood of harm
HARM
label
without
details
USERS / MANIPULATION
PETs that dynamical influence behaviour
misuses of „users"
Actors - GOV, Private parties

Participants

- Frederik Armknecht
Universität Mannheim, DE
- Aurélien Bellet
INRIA – Montpellier, FR
- Robin Berjon
Princeton, US
- Asia Biega
MPI-SP – Bochum, DE
- Paul Comerford
Information Commissioner's Office – Wilmslow, GB
- Ana-Maria Cretu
EPFL – Lausanne, CH
- Emiliano De Cristofaro
University of California – Riverside, US
- Yves-Alexandre de Montjoye
Imperial College London, GB
- Sébastien Gambs
UQAM – Montreal, CA
- Georgi Ganev
University College London, GB
- Patricia Guerra-Balboa
KIT – Karlsruher Institut für Technologie, DE
- Seda F. Gürses
TU Delft, NL
- Johanna Gunawan
Maastricht University, NL
- Bailey Kacsmar
University of Alberta – Edmonton, CA
- Felix Morsbach
KIT – Karlsruher Institut für Technologie, DE
- Lucy Qin
Georgetown University – Washington, DC, US
- Kris Shrishak
Irish Council for Civil Liberties – Dublin, IE
- Thorsten Strufe
KIT – Karlsruher Institut für Technologie, DE
- Hinako Sugiyama
University of California – Irvine, US
- Vanessa Teague
Australian National University – Acton, AU
- Carmela Troncoso
MPI-SP – Bochum, DE
- Michael Veale
University College London, GB
- Rui-Jie Yew
Brown University – Providence, US