

“We are not Future-ready”: Understanding AI Privacy Risks and Existing Mitigation Strategies from the Perspective of AI Developers in Europe

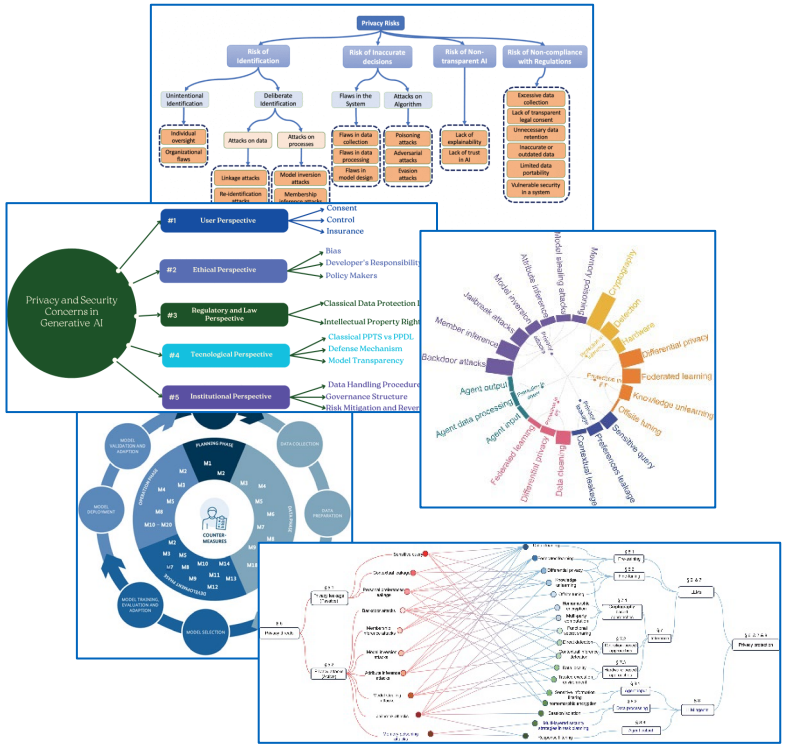
Alexandra Klymenko, **Stephen Meisenbacher**, Patrick Gage Kelley, Sai Teja Peddinti, Kurt Thomas, Florian Matthes

08/11/2025, SOUPS

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Motivation: What do we know?

Many works on systematizing AI privacy risks and mitigations



Understanding AI Privacy Risks and Existing Mitigation Strategies from the Perspective of AI Developers

(In Europe)

Many works on exploring developer privacy perspectives

Understanding user privacy expectations: A software developer's perspective

Awanthika R. Senarath, Nalin Asanka Gamagedara Arachchilage*

Privacy by designers: software developers' privacy mindset

Irit Hadar[1] · Tomer Hasson[1] · Oshrat Ayalon[2] · Eran Toch[2] · Michael Birnhack[3] · Sofia Sherman[1] · Arod Balissa[3]

Evaluating Privacy Perceptions, Experience, and Behavior of Software Development Teams

Maxwell Prybylo, *University of Maine* · Sara Haghighi, *University of Maine* · Sai Teja Peddinti, *Google* · Sepideh Ghanavati, *University of Maine*

"I Never Thought About Securing My Machine Learning Systems": A Study of Security and Privacy Awareness of Machine Learning Practitioners

Franziska Boenisch* franziska.boenisch@aisec.fraunhofer.de Fraunhofer AISEC Germany

Verena Battis* verena.battis@sit.fraunhofer.de Fraunhofer SIT Germany

Nikolas Buchmann nicolas.buchmann@fu-berlin.de Freie Universität Berlin Germany

Maija Poikela maija.poikela@aisec.fraunhofer.de Fraunhofer AISEC Germany

"I Don't Know If We're Doing Good. I Don't Know If We're Doing Bad": Investigating How Practitioners Scope, Motivate, and Conduct Privacy Work When Developing AI Products

Hao-Ping (Hank) Lee[1], Lan Gao[2], Stephanie Yang[2], Jodi Forlizzi[1], Sauvik Das[1]
[1]*Carnegie Mellon University,* [2]*Georgia Institute of Technology*

© sebis

Research Questions

RQ1

What privacy risks do AI developers believe pose the largest risk to the users, developers, and businesses?

RQ2

What mitigation strategies for AI privacy risks have AI developers considered or deployed?

RQ3

What factors do AI developers think will influence the future of privacy in AI?

© sebis

Scope

We focus on the perceptions of **AI developers** in **Europe**.

AI Developers

Spanning the AI development stack:

- including research, data collection, model training and fine-tuning, and deployment or integration into enterprise systems.

Europe

Previous work lacks a focus on European developers

- a significant gap in a region where privacy (and regulation thereof) is at the forefront!

© sebis

Methodology: Semi-Structured Interviews

Interview Participant Demographics and Interview Setup

The Interviews

Key numbers:

- 25 Interviews with AI developers
- Ranging from 1-3 years of experience (n=1) to 20+ (n=4)
- 11 European countries represented
- Interview participants were generally highly educated (Master's and Doctorates)
- Mostly working in the IT sector, as well as finance, consulting, etc.

Interview Guide

- AI Background
- Risk and Mitigation Awareness
- Mitigation Readiness
- Looking Forward



© sebis

Breaking Down AI Privacy Risks and Mitigations

We conducted a tertiary literature review for papers systematizing AI privacy risks and/or mitigations.
We unified the many proposed taxonomies as follows:

Risks

- Data Management
- Data Memorization And Leakage
- Unintended Downstream Usage of Sensitive Prompts and User Metadata
- Adversarial Inference and Inversion
- Misuse via Harmful Applications

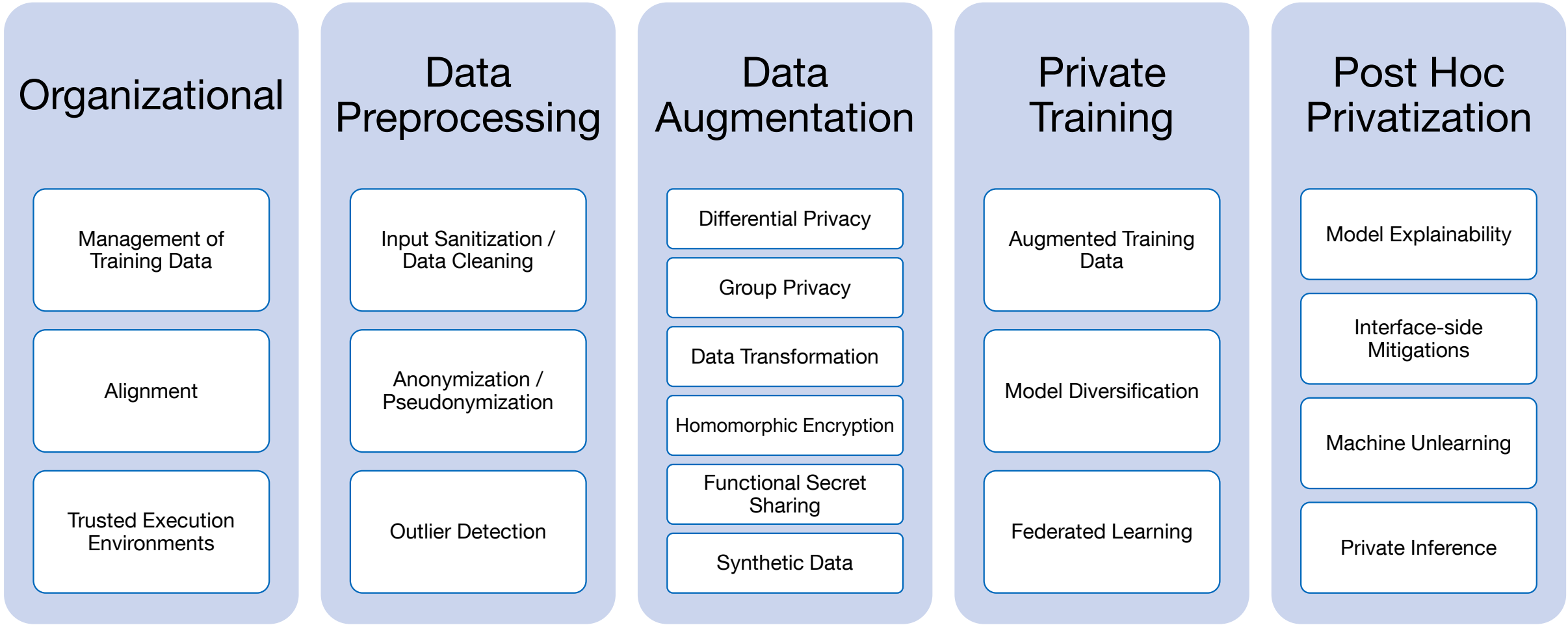
Mitigations

- Organizational
- Data Preprocessing
- Data Augmentation
- Private Training
- Post Hoc Privatization

© sebis

Breaking Down AI Privacy Risks and Mitigations, cont.

We further broke down the mitigation categories into individual mitigation types:



© sebis

Results

What we learned from talking to 25 AI developers in Europe.

Perceptions of Privacy Risks | Mitigation Awareness and Readiness | Looking Forward

Key Findings

Developers pointed to the key **amplification factors** of increased accessibility, capabilities, and use of data that have been enabled by recent AI advances.

> "*I think all the technological [risks] with a bit of effort, we are able to mitigate. Incorporating these technologies **is just a matter of effort**.*"

They cited personal **reasoning patterns** behind perceived risk prevalence, ranging from *technical difficulty* to *risk context*.

Developers perceive the **technical, legal, and societal aspects** of AI privacy risks to be intertwined, with concerns linking privacy risks to broader ethical considerations.

> "*[T]he **societal impact of it is, in the end, what matters**. Technology I would see more as a means to get there. And the same thing for regulation, they should all make sure it has a positive effect on society.*"

© sebis

Results

What we learned from talking to 25 AI developers in Europe.

Perceptions of Privacy Risks | Mitigation Awareness and Readiness | Looking Forward

	No. of times ranked as: #1	#2	#3	#4	#5	Median rank	Average rank	Standard deviation
Misuse via Harmful Applications	7	6	4	3	5	2	2.72	1.51
Data Management	7	5	4	5	4	3	2.76	1.48
Data Memorization and Leakage	7	4	2	6	6	3	3.00	1.61
Unintended Downstream Usage	3	5	6	5	6	3	3.24	1.36
Adversarial Inference and Inversion	1	5	9	6	4	3	3.28	1.10

(Above) AI developers' ranking of AI risks based on their potential harm and urgency to address (highest first). While Misuse *and* Data Management *received the highest scores, there is no clear winner!*

Some risk categories appeared more often at the top…

… and yet, overall, all ended up "somewhere in the middle"

© sebis

Results

What we learned from talking to 25 AI developers in Europe.

Perceptions of Privacy Risks | **Mitigation Awareness and Readiness** | Looking Forward

Key Findings

AI privacy risk mitigation strategies received **varying degrees of attention** in the interviews.

Some mitigations were “standard practice” or the “gold standard”, while others were **familiar by name only**.

Some **contentious (but well-known) mitigations**, e.g., Differential Privacy, Synthetic Data, and “last resort” options (like Unlearning).

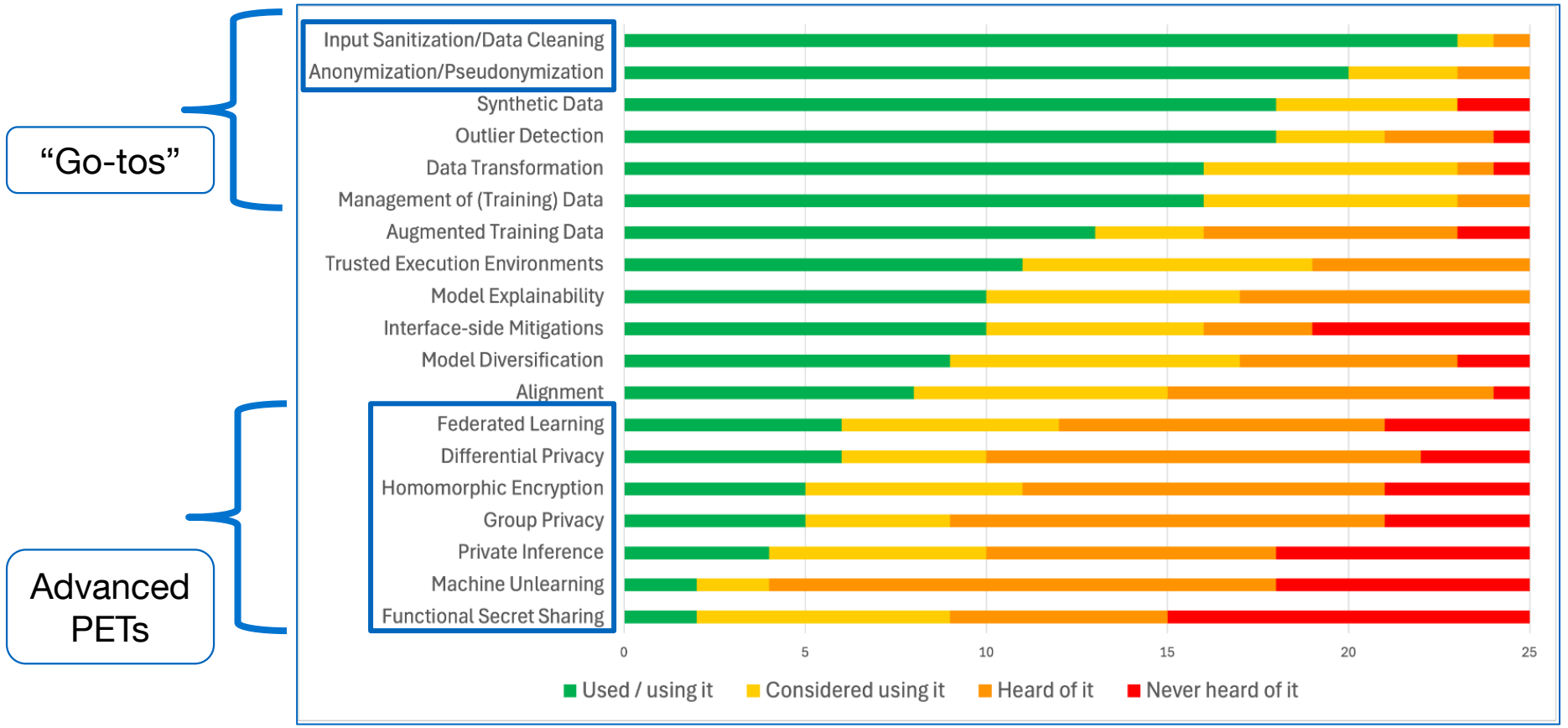
> “*The biggest impact you can make is **reduce the number of bad decisions** that someone in an organization makes.*”

© sebis

Results

What we learned from talking to 25 AI developers in Europe.

Perceptions of Privacy Risks | **Mitigation Awareness and Readiness** | Looking Forward



(Left) Participant self-ranking of their mitigation familiarity and experience in adopting these techniques. More "traditional" techniques (and synthetic data) were reported to be used the most often, while more advanced PETs often received "heard of it" or "never heard of it".

© sebis

Results

What we learned from talking to 25 AI developers in Europe.

Perceptions of Privacy Risks | Mitigation Awareness and Readiness | Looking Forward

Key Findings

The interviewed developers indicated a number of **key stakeholders in AI privacy**, particularly the government and AI providers.

Need **clearer guidelines** and more **usable mitigation tools**, with the goal of increased awareness, transparency, and collaboration.

Key Stakeholders

- Government and Regulatory Bodies
- AI Providers
- End Users
- Academia / Research

Opportunities

- Regulations, Standards, and Guidelines
- Tools and Testing Systems
- Education and Awareness
- Transparency
- Interdisciplinary Collaboration

© sebis

Results

What we learned from talking to 25 AI developers in Europe.

Perceptions of Privacy Risks | Mitigation Awareness and Readiness | **Looking Forward**

> "***We are not future-ready***. *I don't know if anyone is future-ready, and or even the short[-term] future-ready.* ***We are just trying to save the day.***"

© sebis

Which AI privacy risks are most pressing?

- 25 ranking exercises, 23 unique orderings!
- Although certain risks are generally perceived as more urgent than others, this is not uniformly the case.

© sebis

Discussion

RQ1: Navigating the Sea of Privacy Risks

Which AI privacy risks are most pressing?

- 25 ranking exercises, 23 unique orderings!
- Although certain risks are generally perceived as more urgent than others, this is not uniformly the case.

AI privacy risk perception is highly personal…

- … and interpretive!
- Reasons given were generally based on some personal, self-guided reasoning pattern (e.g., “*If X is easier to realize than Y, then X is more of a threat.*”)

© sebis

Discussion

RQ1: Navigating the Sea of Privacy Risks

Which AI privacy risks are most pressing?

- 25 ranking exercises, 23 unique orderings!
- Although certain risks are generally perceived as more urgent than others, this is not uniformly the case.

AI privacy risk perception is highly personal…

- … and interpretive!
- Reasons given were generally based on some personal, self-guided reasoning pattern (e.g., "*If X is easier to realize than Y, then X is more of a threat.*")

What does this mean for regulation?

- Regulators looking to the industry for answers might be left without clear / uniform answers.
- Leading to premature regulatory enforcement of AI?

© sebis

Discussion

RQ2: Mitigations Abound, but Adoption is Lagging

Generally high awareness of mitigations…

- With some receiving considerable attention over others – especially those “popularized” by recent publications, regulation, or even media.
- Possibly implying generally increasing awareness of privacy risk mitigations among (AI) industry professionals.

© sebis

Discussion

RQ2: Mitigations Abound, but Adoption is Lagging

Generally high awareness of mitigations…

- With some receiving considerable attention over others – especially those "popularized" by recent publications, regulation, or even media.
- Possibly implying generally increasing awareness of privacy risk mitigations among (AI) industry professionals.

Some mitigations are "go-tos"

- "Anonymization" was a quite popular topic.
- While other "advanced" privacy technologies were often perceived to have significant shortcomings, despite their more rigorous privacy notions.

© sebis

Discussion

RQ2: Mitigations Abound, but Adoption is Lagging

Generally high awareness of mitigations…

- With some receiving considerable attention over others – especially those "popularized" by recent publications, regulation, or even media.
- Possibly implying generally increasing awareness of privacy risk mitigations among (AI) industry professionals.

Some mitigations are "go-tos"

- "Anonymization" was a quite popular topic.
- While other "advanced" privacy technologies were often perceived to have significant shortcomings, despite their more rigorous privacy notions.

How to address the divide?

- Important to distinguish true practical limitations from resistance to change / more complex mitigations.
- *Adoption* rather than pure awareness – clear call to researchers and regulators alike.

© sebis

AI providers “hold the keys”

- But with a general feeling that nothing will be done without a clear demand from enterprise customers and end users to make privacy a priority.
- This is an optimistic view! Consumer focus on privacy over functionality will likely not come without intervention.

© sebis

Discussion

RQ3: Becoming Future-ready

AI providers “hold the keys”

- But with a general feeling that nothing will be done without a clear demand from enterprise customers and end users to make privacy a priority.
- This is an optimistic view! Consumer focus on privacy over functionality will likely not come without intervention.

The role of researchers and governmental bodies

- We echo previous works in stressing the importance of including privacy in AI education.
- Can be boosted with the support of governmental and other supervisory authorities.

© sebis

Conclusion and Outlook

We conduct a study with 25 AI developers across Europe, looking at their perceptions of AI privacy risks and currently proposed mitigation strategies.

Takeaways and Future Work

- Researchers should strive for greater clarity on how AI privacy risks are understood in practice.
- Focus on clarity of what AI privacy risks entail, particularly for non-technical audiences.
- Make privacy risk mitigation strategies accessible to AI developers (which requires collaboration!).

Limitations

- Geographic scope – Europe is a very diverse place!
- Skew towards more educated participants / those with experience in research.
- Clear bias towards male participants (88%).

"We are not Future-ready": Understanding AI Privacy Risks and Existing Mitigation Strategies from the Perspective of AI Developers in Europe

Alexandra Klymenko* *Technical University of Munich*
Stephen Meisenbacher* *Technical University of Munich*
Patrick Gage Kelley *Google*
Sai Teja Peddinti *Google*
Kurt Thomas *Google*
Florian Matthes *Technical University of Munich*

Check out our paper for (many) more details and insights not covered in this presentation. We are happy to discuss!

Chair of Software Engineering for Business Information Systems
Department of Computer Science
School of Computation, Information and Technology
Technical University of Munich

Investigating User Perspectives on Differentially Private Text Privatization

Stephen Meisenbacher, Alexandra Klymenko, Alexander Karpp, and Florian Matthes
{stephen.meisenbacher,alexandra.klymenko,a.karpp,matthes}@tum.de

When would YOU share your personal text data under Differential Privacy?

Shameless plug: also stop by our poster at tonight's poster session!



Stephen Meisenbacher
stephen.meisenbacher@tum.de

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

+49.89.289.17132
matthes@in.tum.de
wwwmatthes.in.tum.de