UNIVERSITÄT OSNABRÜCK

Prädiktive Privatheit. Datenschutz im Zeitalter der KI

Jahrestagung Forum Privatheit

18. November 2021

Rainer Mühlhoff
Professor für Ethik der Künstlichen Intelligenz
Institut für Kognitionswissenschaft
Universität Osnabrück

@RainerMuehlhoff
https://RainerMuehlhoff.de
rainer.muehlhoff@uni-osnabrueck.de

Privatheit vs. Datenschutz

- Datenschutz ist nicht auf Schutz der Privatsphäre reduzibel.

- Schutz der Privatsphäre ist *ein Teilaspekt* eines gelingenden Datenschutzes.

„Es geht nicht um Privatsphäre […], es geht darum, eine Technik sozial beherrschbar zu machen.“

(Wilhelm Steinmüller, 2009)

Datenschutz: Dominante Angriffsszenarien

	Typ 1: Überwachung, Intrusion	**Typ 2: Re-Identifikation**	**Typ 3: Vorhersage, social sorting**
Virulent seit	1970 ff.	1990 ff.	**2010 ff.**

Datenschutz: Dominante Angriffsszenarien

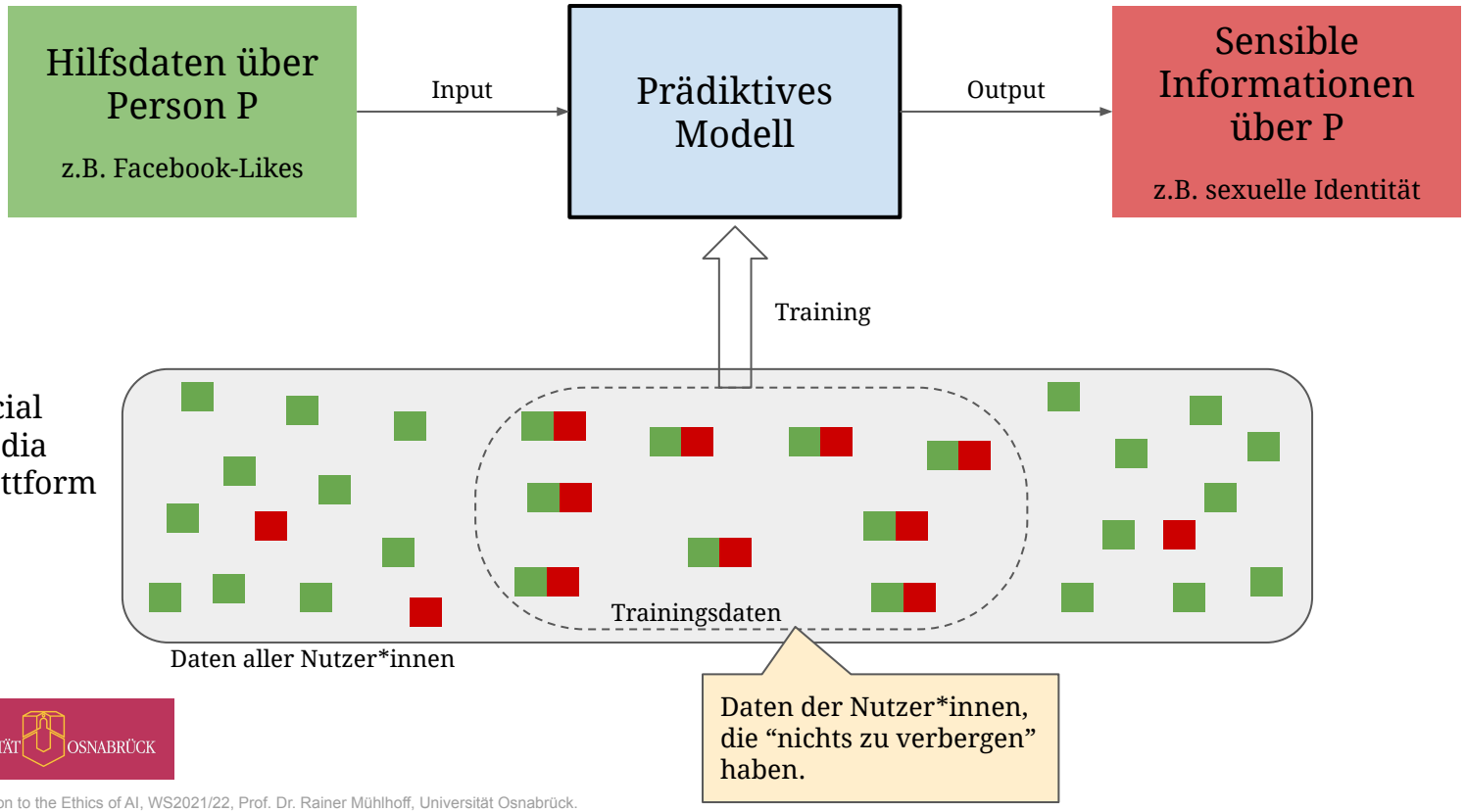
	Typ 1: Überwachung, Intrusion	**Typ 2: Re-Identifikation**	**Typ 3: Vorhersage, social sorting**
Virulent seit	1970 ff.	1990 ff.	**2010 ff.**
Angriffsziel	personenbezogene Daten	Anonymität in großen Datensätzen	**Gleichheit der Behandlung, Fairness**

Datenschutz: Dominante Angriffsszenarien

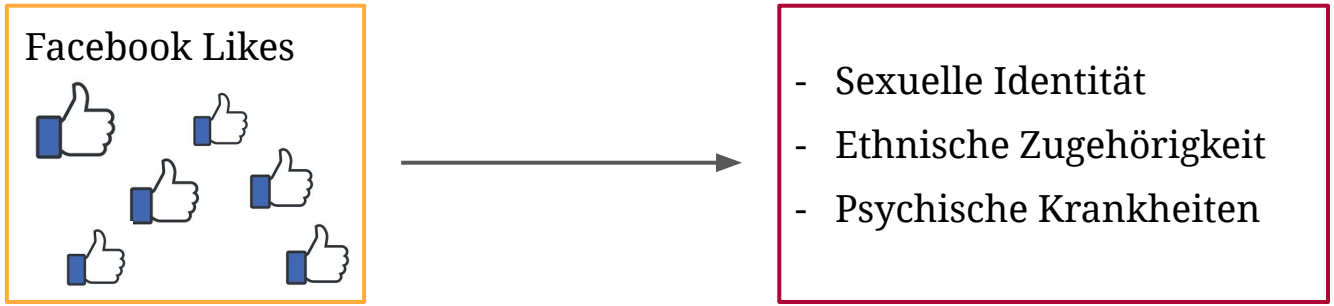
	Typ 1: Überwachung, Intrusion	**Typ 2: Re-Identifikation**	**Typ 3: Vorhersage, social sorting**
Virulent seit	1970 ff.	1990 ff.	**2010 ff.**
Angriffsziel	personenbezogene Daten	Anonymität in großen Datensätzen	**Gleichheit der Behandlung, Fairness**
Schutz	Datensicherheit	Differential Privacy, Federated Machine Learning	**Predictive Privacy**

1. Prädiktive Analytik

Prädiktive Analytik – Funktionsweise



Beispiele



Kosinski, M, Stillwell, D, and Graepel, T. 2013. “Private traits and attributes are predictable from digital records of human behavior”. *Proceedings of the National Academy of Sciences* 110.15, 5802–5805.

Merchant, R M, Asch, D A, Crutchley, P, Ungar, L H, Guntuku, S C, Eichstaedt, J C, Hill, S, Padrez, K, Smith, R J, and Schwartz, H A. 2019. “Evaluating the predictability of medical conditions from social media posts”. *PLOS ONE* 14.6, e0215476.

Reilly, M. 2017. “Is Facebook Targeting Ads at Sad Teens?” *MIT Technology Review*.

Beispiele

Browser-Verlauf

- Versicherungsrisiken
- Kreditwürdigkeit
- Bildungsgrad

Eubanks, V. 2017. *Automating inequality: how high-tech tools profile, police, and punish the poor*. First Edition New York, NY.

Hurley, M and Adebayo, J. 2017. “Credit scoring in the era of big data”. *Yale Journal of Law and Technology* 18.1, 5.

O’Dwyer, R. 2018. “Are You Creditworthy? The Algorithm Will Decide.” *Undark Magazine*.

O’Neil, C. 2016. *Weapons of math destruction: how big data increases inequality and threatens democracy*. First edition New York.

2. Prädiktive Privatheit

Das Konzept

“Die prädiktive Privatheit einer Person wird verletzt, wenn sensible Informationen ohne ihr Wissen und gegen ihren Willen über sie vorhergesagt werden, …

… und zwar in einer Weise, die Auswirkungen auf ihre Chancen und ihr Wohlergehen hat.”

Mühlhoff, Rainer. “Predictive Privacy: Towards an Applied Ethics of Data Analytics.” *Ethics and Information Technology,* 2021. https://doi.org/10.1007/s10676-021-09606-x.
https://predictiveprivacy.org

Prädiktive Privatsphäreverlestzung

Aspekt 1:
Sensible Informationen werden aus scheinbar weniger sensiblen Informationen abgeleitet.

Mühlhoff, Rainer. “Predictive Privacy: Towards an Applied Ethics of Data Analytics.” *Ethics and Information Technology,* 2021. https://doi.org/10.1007/s10676-021-09606-x.
https://predictiveprivacy.org

Prädiktive Privatsphäreverlestzung

Aspekt 2:
Sensible Informationen über das betroffene Individuum werden anhand von Daten abgeschätzt, die viele andere Individuen über sich preisgegeben haben.

Mühlhoff, Rainer. “Predictive Privacy: Towards an Applied Ethics of Data Analytics.” *Ethics and Information Technology,* 2021. https://doi.org/10.1007/s10676-021-09606-x.
https://predictiveprivacy.org

- Die Daten *anderer* haben Auswirkungen auf einen selbst.
- Die eigenen Daten haben Auswirkungen auf andere.
- Das gilt auch für anonymisierte Daten.

Datenschutz ist keine private Entscheidung

3. Konsequenzen

Einige aktuelle Trends im “Datenschutz” *ermöglichen* prädiktive Analytik (paradoxerweise)

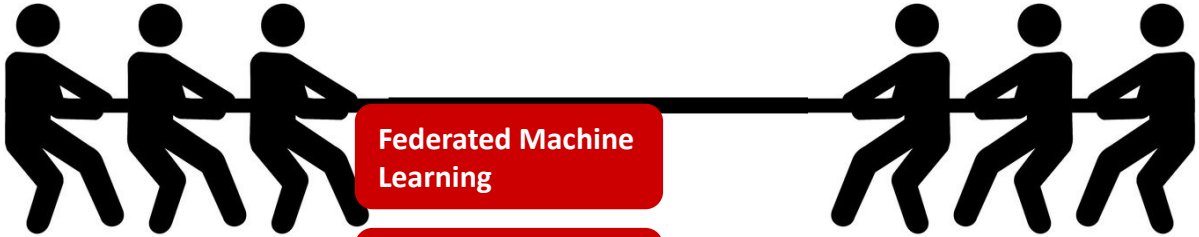
1. Fokussierung auf individuelle Einwilligung und “privacy self-management”.
2. PR-wirksame Betonung der technischen Garantie *individueller* Anonymität der Nutzer*innen.

Vgl. Kröger; Lutz; Ullrich (2021). The myth of individual control: Mapping the limitations of privacy self-management. In *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3881776

Informierte Einwilligung

- suggeriert fälschlicherweise, dass die Entscheidung nur von individueller Tragweite ist;
- gibt großen Unternehmen in der Praxis freie Hand zur Erhebung und Verwendung Verhaltensdaten.

“Differential Privacy” *ermöglicht* prädiktive Analytik



Differential Privacy

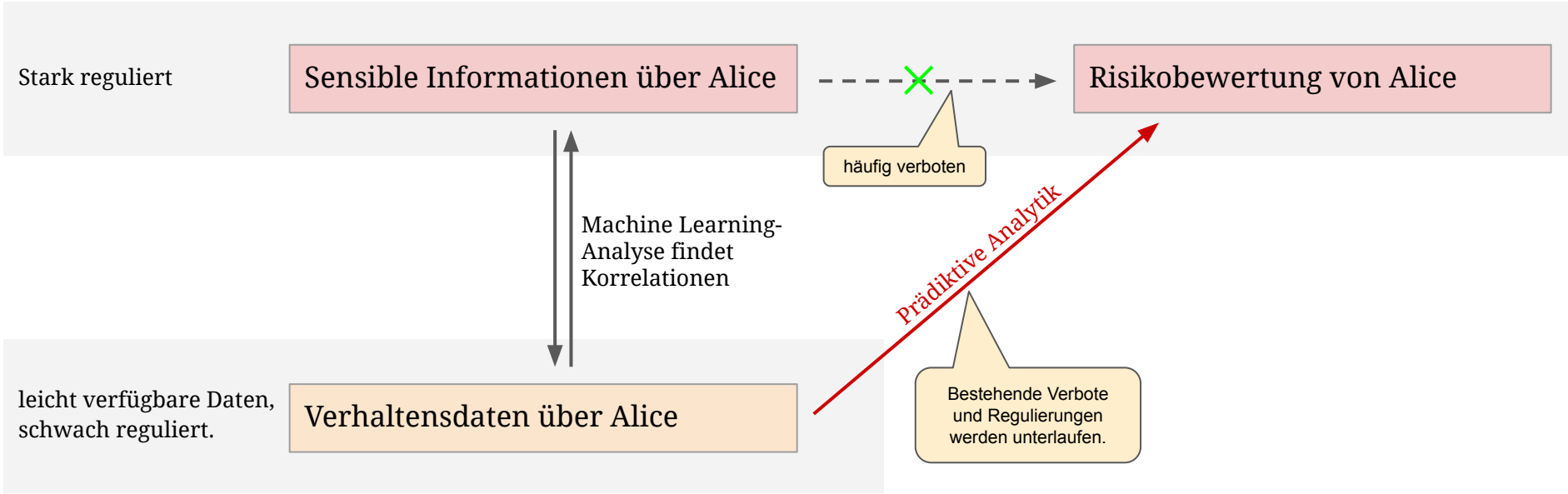
Über die Individuen in den Trainingsdaten eines prädiktiven Modells werden keine zusätzlichen Informationen ableitbar dadurch, dass das Individuum in den Trainingsdaten ist.

- individualistischer Standpunkt
- *dazu gemacht,* prädiktive Analytik unter dem Label der “Privatheit” zu ermöglichen

Predictive Privacy

Ein prädiktives Modell verletzt die prädiktive Privatheit beliebiger Individuen, wenn es die Abschätzung von Informationen ermöglicht, die man nicht preisgeben möchte.

Regulatorischer Handlungsbedarf



Vorschlag: Zweigleisige Vorgehensweise

1. Prädiktive Informationen rechtlich wie personenbezogene Daten stellen

   → grundsätzliches Verarbeitungsverbot + Rechtsgrundlage/Erlaubnisvorbehalt

2. In definierten Anwendungsbereichen die Herstellung prädiktiver Risiko-Modelle kategorisch verbieten.

   → z.B. Haftentscheidungen, Bewerbungsverfahren, Krankenversicherungen, Finanzindustrie, …

Vielen Dank!

Rainer Mühlhoff
Professor für Ethik der Künstlichen Intelligenz

Institut für Kognitionswissenschaft
Universität Osnabrück

https://RainerMuehlhoff.de | rainer.muehlhoff@uni-osnabrueck.de | @RainerMuehlhoff

https://predictiveprivacy.org

UNIVERSITÄT OSNABRÜCK

Datenschutz: Dominante Angriffsszenarien

	Typ 1: Datenklau/Intrusion	**Typ 2: Re-Identifikation**	**Typ 3: Vorhersage**
Virulent seit	1970 ff.	1990 ff.	**2010 ff.**
Beschreibung	Trotz sicherer Speicherung und Übertragung gelingt Unbefugten ein Zugriff auf sensible Daten.	Trotz anonymisierter Datenverarbeitung sind Individuen in großen Datensätzen identifizierbar.	Daten vieler anderer User werden dafür verwendet, Informationen über eine bestimmte Nutzer*in zu schätzen.
Angriffsziel	sensible Datenbestände	Identität / Anonymität	Gleichheit der Behandlung, Fairness
Methode	- Datenklau (Hacker, NSA) - Datenlecks	Re-Identifikation durch - Verknüpfung mit Hilfsdaten, - statistische Angriffe	- "Pattern Matching" - Machine Learning, KI
Schutz	Datensicherheit	Differential Privacy, Federated Machine Learning	**Predictive Privacy**