Wie kommen wir an die Daten?
KI und der Datenschutz

Hannes Federrath · Joshua Stock
Sicherheit in verteilten Systemen (SVS)
http://svs.informatik.uni-hamburg.de

Der Arbeitsbereich Sicherheit in Verteilten Systemen (SVS)

- Unsere Forschungsthemen (Auswahl)
  - IT-Sicherheitsmanagement, und -Grundschutz, ISO 27001
  - Privacy im Internet, Schutz vor Beobachtung, IT-Forensik
  - Sichere und datenschutzfreundliche Vernetzung von Fahrzeugen
  - Sicherheit und Datenschutz in mobilen Systemen

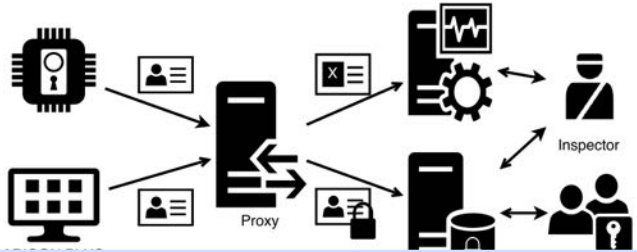
- Beiträge und (interdisziplinäre) Ergebnisse
  - Begleitung von Gesetzgebungsverfahren aus technischer Sicht
  - Erforschung des Spannungsfeldes von Freiheit und Sicherheit
  - Technische Lösungen zum Grundrechtsschutz
  - Informatik als gesellschaftliche Aufgabe

- Weitere Informationen
  - https://svs.informatik.uni-hamburg.de

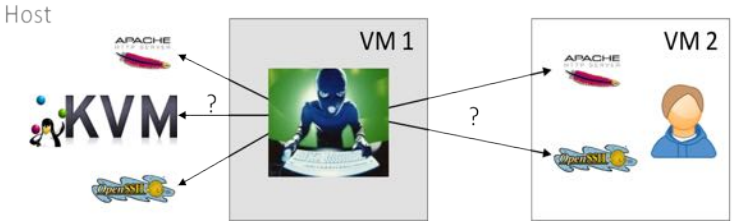
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Forschung: Datenschutz und Datensicherheit (Auswahl)

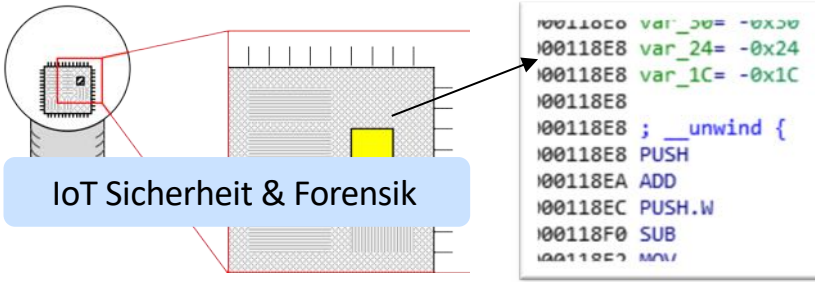
Privatsphäreangriffe und Fragen des technischen Datenschutzes in der KI



Sicherheit und Privatsphäre in Anwendungen, Datenschutzfreundliche Angriffserkennung



Sicherheit und Privatsphäre in der Cloud



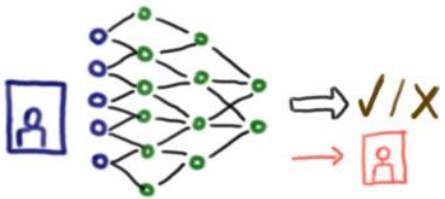
IoT Sicherheit & Forensik

Aufgaben von SVS in KI-SIGS (AP220)

- Anonymisieren und Pseudonymisieren von Daten innerhalb der technischen Plattform
  - Programmierbibliothek für datenschutzgerechte Datenaufbereitung und -analyse
- Rechtsgrundlagen zur Datenverarbeitung mit KI-Systemen
  - Wie muss rechtswirksame Einwilligung gemäß DSGVO gestaltet sein?
- Datenschutz-Folgenabschätzung (DSFA) nach Art. 35 DSGVO
  - Spezielle Datenschutzrisiken von KI-Systemen beachten
  - Entwicklung einer DSFA-Methodik für KI-Systeme
- Datenbereitstellungsverträge

Thesen zum Datenschutz im KI-Kontext

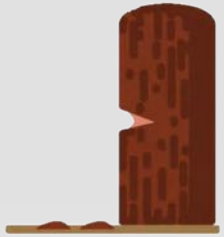
- Training von KI-Algorithmen benötigt große Mengen an Daten
  - oftmals personenbezogene Daten
  - medizinischer Bereich: sensible Daten (Art. 9 DSGVO)
- Zensieren von sensiblen Daten reicht nicht aus
  - von KI reproduzierbar
- KI-Modelle lernen mehr als für die jeweilige Aufgabe notwendig
  - Repräsentation der Daten bleibt in KI-Modell bestehen
    - auch von einzelnen Datenpunkten



Folge: Privatsphäre-Angriffe sind u.U. möglich. Die Abschätzung der Risiken *vor* der Einführung eines (KI)-Verfahrens ist notwendig.

Abbildungen aus dem Poster

Regulatorische Plattform



Softwareentwickler

Software Code basierend auf Anforderungen

Endprodukt

Trainingsdaten

Bibliotheken

Tools fürs Training

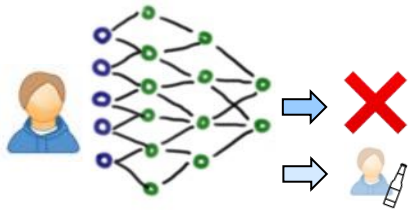
+

Software Code generiert durch Software

Endprodukt

Beispiel 1: *Membership Inference* Angriff [SS+17]

- Patient Max: alkoholkrank
- Einwilligung in Suchtklinik zu Verarbeitung seiner anonymisierten Patientendaten
  - Training eines KI-Modells der Firma *AlcScore*

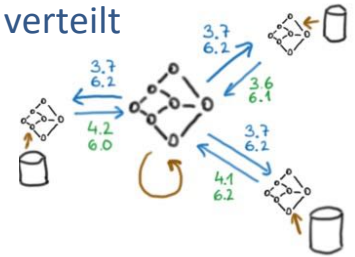


- Idee von *AlcScore*:
  - Anhand von Patientendaten Risikofaktoren extrahieren
  - Eingabe in KI-Modell: Anonymisierte Patientendaten
  - Ausgabe: Risikoscore Neigung zu Alkoholismus

- Jahre später: Max möchte Lebensversicherung abschließen und muss Gesundheitsdaten angeben
  - Versicherung hat Modell von *AlcScore* akquiriert
  - Kann anhand Max‘ Daten nicht nur Score berechnen
  - ↯ durch *Membership Inference* [SS+17]: Max war Teil der *AlcScore*-Trainingsdaten
  - Diagnostizierte Sucht –> Versicherung lehnt Vertrag mit Max ab

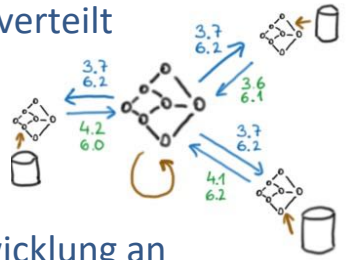
Beispiel 2: *Federated Learning* Rekonstruktionsangriff [ZH20]

- *Federated Learning* (FL): Trainingsdaten für Modell sind auf mehrere Standorte verteilt
  - Vorteil: Daten müssen nicht ausgetauscht werden
  - Zentraler Server koordiniert
    - Jeder Standort trainiert lokales Modell
    - Schickt aktualisierte Modellparameter an Server
    - Server aggregiert Parameter, aktualisiert und verteilt globales Modell

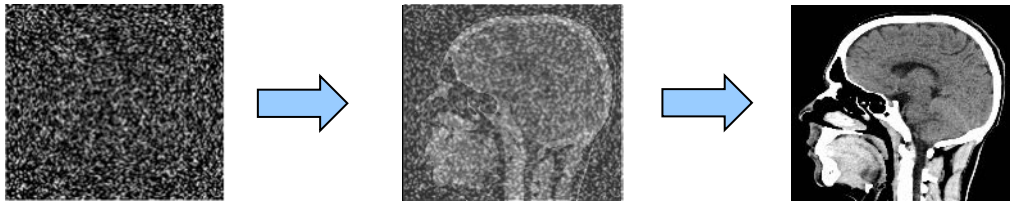


Beispiel 2: *Federated Learning* Rekonstruktionsangriff [ZH20]

- *Federated Learning* (FL): Trainingsdaten für Modell sind auf mehrere Standorte verteilt
  - Vorteil: Daten müssen nicht ausgetauscht werden
  - Zentraler Server koordiniert, aggregiert und verteilt globales Modell



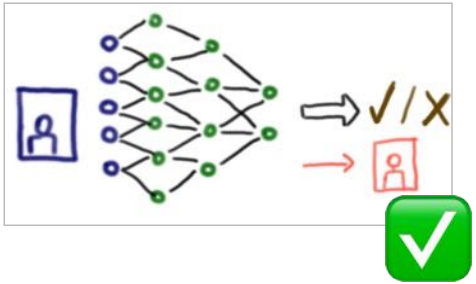
- Beispiel: Gesundheitsdienstleister bietet FL-Dienst für gemeinsame Modellentwicklung an
  - Hirntumorerkennung auf CT-Bildern
  - Kliniken nehmen teil: lediglich Modellupdates, aber keine CT-Bilder werden übertragen
  - Dienstleister kann aus Modellupdates einzelner Klinik Bilder rekonstruieren [ZH20]
    - Start bei zufälligem Rauschen
    - iterativer Prozess, bis Modellparameter 1:1 übereinstimmen



Lösungswege *Privacy Preserving Machine Learning* (PPML)

- Anonymisierung und Pseudonymisierung
  - Nicht trivial; bedingt wirksam gegen KI-Privatsphäreangriffe
- Differential Privacy [DM+06]
  - Kontextabhängige, gezielte Manipulation von Algorithmen oder Daten
  - Minimiert Einfluss einzelner Datenpunkte –> wirksam gegen *Membership Inference* und Rekonstruktionsangriffe [AC+16]
- Secure Multiparty Computation (SMPC)
  - Kryptographische Lösung für verteiltes Training
- Homomorphe Verschlüsselung (HE)
  - Ermöglicht Training und Inferenz auf verschlüsselten Daten

SMPC und HE: rechenintensiv und (noch) nicht für Masseneinsatz tauglich [CP21]



Privacy by Design
Art. 25 und 32 DSGVO

Privacy by Design

Stand der Technik

Kosten

ROSI

Art, Umfang, Umstände, Zwecke

Privacy by Default

- Beschränkung auf Verarbeitungszweck
- Beschränkung des Umfangs
- Beschränkung auf Speicherfristen
- Beschränkung der Zugänglichkeit

Risiko-management

Technisch-organisatorische Maßnahmen

Datenschutzgrundsätze

Betroffenenrechte

Rechtmäßigkeit, Transparenz

Zweck-bindung

Rechenschafts-pflicht

Richtigkeit

Speicher-begrenzung

Daten-minimierung

Daten-sicherheit

- Auskunft
- Berichtigung
- Löschung
- Einschränkung der Verarbeitung
- Datenübertragbarkeit
- Widerspruchsrecht

- Verbot mit Erlaubnisvorbehalt oder Einwilligung
- Ersterhebung bei betroffener Person
- Gute Dokumentation
- ...

Privacy Policy

- Anonymität
- Unbeobachtbarkeit
- Unverkettbarkeit

- Pseudonymisierung
- Verschlüsselung
- Vertraulichkeit
- Integrität
- Verfügbarkeit
- Belastbarkeit
- Wiederherstellbarkeit
- Notfallmanagement
- Überprüfbarkeit
- Bewertung
- Evaluierung

PIMS

Bereitschaft zur datenschutz-freundlichen Gestaltung

Big Data

Algorithmen

- Enforce
- Inform
- Control
- Demonstrate

- Minimise
- Separate
- Aggregate
- Perturbate
- Hide

Initiierung

Si-Konzept

Umsetzung

Erhaltung

Kontext

Bewertung

Gew-Ziele

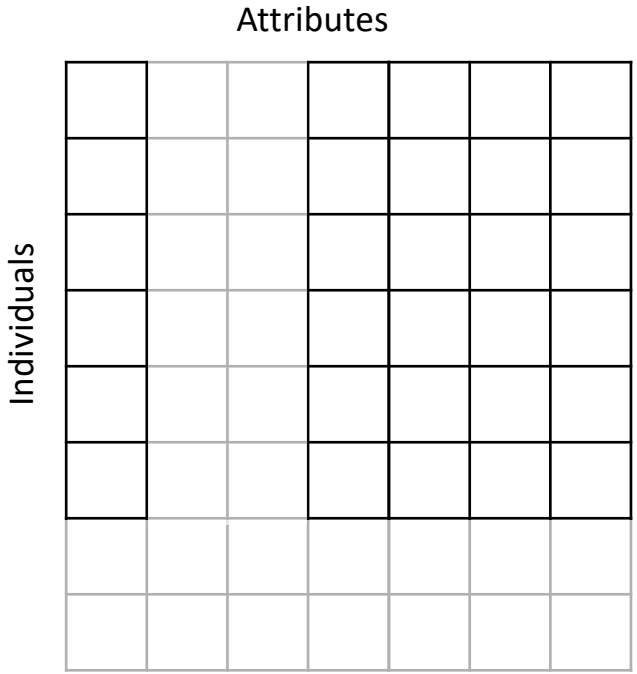
Soll-Ist-An

Feedback

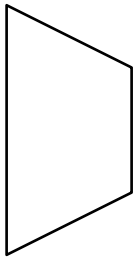
Bereitschaft

Privacy design strategies

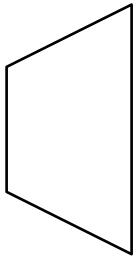
in Anlehnung an Hoepman, 2014



minimise

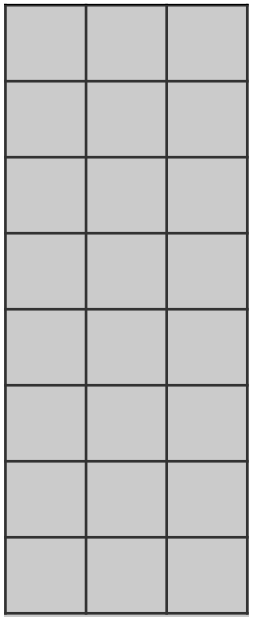


aggregate



perturbate

separate



hide

Privacy by Design (Art. 25 DSGVO) und Sicherheit der Verarbeitung (Art. 32 DSGVO)

Minimise: Nur notwendige Daten speichern und verarbeiten	Anonymisierung, Pseudonymisierung
Separate: Daten verteilt verarbeiten und speichern	Federated Learning, SMPC
Aggregate: Daten auf das notwendige Maß zusammenfassen	Federated Learning
Perturbate: Daten durch zufällige Störungen ungenau machen	Differential Privacy
Hide: Daten nicht in offener Form speichern	Homomorphic Encryption

Privacy by Design

Stand der Technik

Kosten

ROSI

Art, Umfang, Umstände, Zwecke

Privacy by Default

- Beschränkung auf Verarbeitungszweck
- Beschränkung des Umfangs
- Beschränkung auf Speicherfristen
- Beschränkung der Zugänglichkeit

Risiko-management

Technisch-organisatorische Maßnahmen

Datenschutzgrundsätze

Betroffenenrechte

Rechtmäßigkeit, Transparenz

Zweck-bindung

Rechenschafts-pflicht

Richtigkeit

Speicher-begrenzung

Daten-minimierung

Daten-sicherheit

- Auskunft
- Berichtigung
- Löschung
- Einschränkung der Verarbeitung
- Datenübertragbarkeit
- Widerspruchsrecht

- Verbot mit Erlaubnisvorbehalt oder Einwilligung
- Ersterhebung bei betroffener Person
- Gute Dokumentation
- ...

Privacy Policy

- Anonymität
- Unbeobachtbarkeit
- Unverkettbarkeit

- Pseudonymisierung
- Verschlüsselung
- Vertraulichkeit
- Integrität
- Verfügbarkeit
- Belastbarkeit
- Wiederherstellbarkeit
- Notfallmanagement
- Überprüfbarkeit
- Bewertung
- Evaluierung

PIMS

Bereitschaft

Kontext

Bewertung

Gew-Ziele

Soll-Ist-An

Feedback

Bereitschaft zur datenschutz-freundlichen Gestaltung

Big Data

Algorithmen

- Enforce
- Inform
- Control
- Demonstrate

- Minimise
- Separate
- Aggregate
- Perturbate
- Hide

Initiierung

Si-Konzept

Umsetzung

Erhaltung

Zum Schluss: KI-Einsatz im Rechtlichen Kontext

- Rechtsgrundlage für Verarbeitung, Datenschutz-Folgenabschätzung -> KI-SIGS, AP 220
- Automatisierte Entscheidungen (Art. 22 DSGVO)
  - Rechtliche Autonomie, KI-System als juristische Person?
- Sind KI-Modelle personenbezogene Daten?
  - Privatsphäre-Angriffe ermöglichen Rekonstruktion
- Umsetzung des Rechts auf Vergessenwerden (Art. 17 DSGVO) und Recht auf Berichtigung (Art. 16)
  - Neu-Trainieren von KI-Modellen?
- Gewährleistung von Transparenz der Datenverarbeitung, Vermeidung von Bias


Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Maschinelles Lernen aus Sicht der Privatsphäre

Joshua Stock, Tom Petersen, Hannes Federrath

18. Oktober 2021

1 Einleitung

2 Grundlagen

2.1 Lerntypen

inf.uni-hamburg.de

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DEPARTMENT OF INFORMATICS
SECURITY AND PRIVACY

HOME COURSES THESES RESEARCH PEOPLE SERVICE

UHH → MIN-Fakultät → Fachbereich Informatik → Einrichtungen → Arbeitsbereiche → Security and Privacy → Home

WORKING GROUP ON «SECURITY AND PRIVACY»

Security and Privacy

Information systems become more and more important in critical infrastructures, while the Internet has evolved to a critical infrastructure itself. The secure operation of these infrastructures is vital and their failure can have severe impacts up to the loss of human lives.
Security refers to the fact that protection goals are achieved in the presence of malicious attacks and system failures. Typical security goals can be confidentiality, integrity, accountability, and availability. Security and privacy in information systems addresses both technical and organizational aspects, such as building and establishing security concepts and security infrastructures as well as risk analysis and risk management.
Privacy can be a conflicting goal to security, but they can also benefit from each other. Hence, it is necessary to balance both when developing secure information systems.

Prof. Dr. Hannes Federrath
Fachbereich Informatik
Universität Hamburg
Vogt-Kölln-Straße 30
D-22527 Hamburg

Telefon +49 40 42883 2358

federrath@informatik.uni-hamburg.de

https://svs.informatik.uni-hamburg.de

Quellen

[AC+16] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (pp. 308-318).

[CP21] Cabrero-Holgueras, J., & Pastrana, S. (2021). SoK: Privacy-Preserving Computation Techniques for Deep Learning. Proceedings on Privacy Enhancing Technologies, 2021(4), 139-162.

[DM+06] Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006, March). Calibrating noise to sensitivity in private data analysis. In Theory of cryptography conference (pp. 265-284). Springer, Berlin, Heidelberg.

[SS+17] Shokri, R., Stronati, M., Song, C., & Shmatikov, V. (2017, May). Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 3-18). IEEE.

[ZH20] Zhu, L., & Han, S. (2020). Deep leakage from gradients. In Federated learning (pp. 17-31). Springer, Cham.