Privacy by Blockchain Design

Werkzeuge zur Messung von Anonymität

insbes. Anonymity Assessment

STIFTUNG DATENSCHUTZ

Das Team

Michael Kolain
Rechtswissenschaftler

Vorstand - Robotics and AI Legal Society (RAILS)
ehem. Koordinator am FÖV Speyer

kolain@ai-laws.org

Christian Grafenauer
Computerwissenschaftler

Privacy Engineer - TechGDPR
Vertreter des Verbraucherinteresses - DIN

Grafenauer@protectivecircle.com

Anonymity Assessment

April 2020

DIN SPEC 4997

DIN

Privacy by Blockchain Design: Ein standardisiertes Verfahren für die Verarbeitung personenbezogener Daten mittels Blockchain-Technologie; Text Englisch

Privacy by Blockchain Design: A standardised model for processing personal data using blockchain technology;
Text in English

3.1
Anonymity assessment
Reidentification assessment
a procedure to evaluate the residual risks of identifying a natural person with data that is claimed to be anonymous

ein Verfahren, um die Restrisiken einer Identifizierung natürlicher Personen im Rahmen konkreter Datenverarbeitungsvorgänge zu bewerten

Kontext

RAILS.
Robotics & AI Law Society

TUM MSRM
Geriatronics

Anonymity Assessment – A Universal Tool for Measuring Anonymity of Data Sets Under the GDPR with a Special Focus on Smart Robotics

Rutgers University Computer & Technology Law Journal, Vol. 48, No. 2, 2022

30 Pages • Posted: 11 Feb 2022

Michael Kolain
German Research Institute for Public Administration (FÖV Speyer)

Christian Grafenauer
affiliation not provided to SSRN

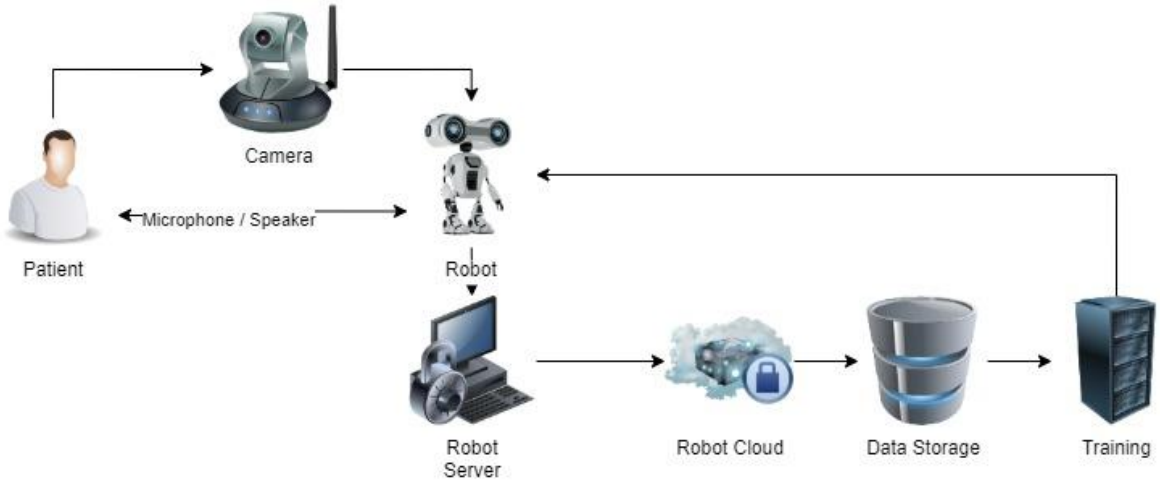
Martin Ebers
Humboldt University of Berlin - Faculty of Law; University of Tartu, School of Law

Date Written: November 24, 2021

Abstract

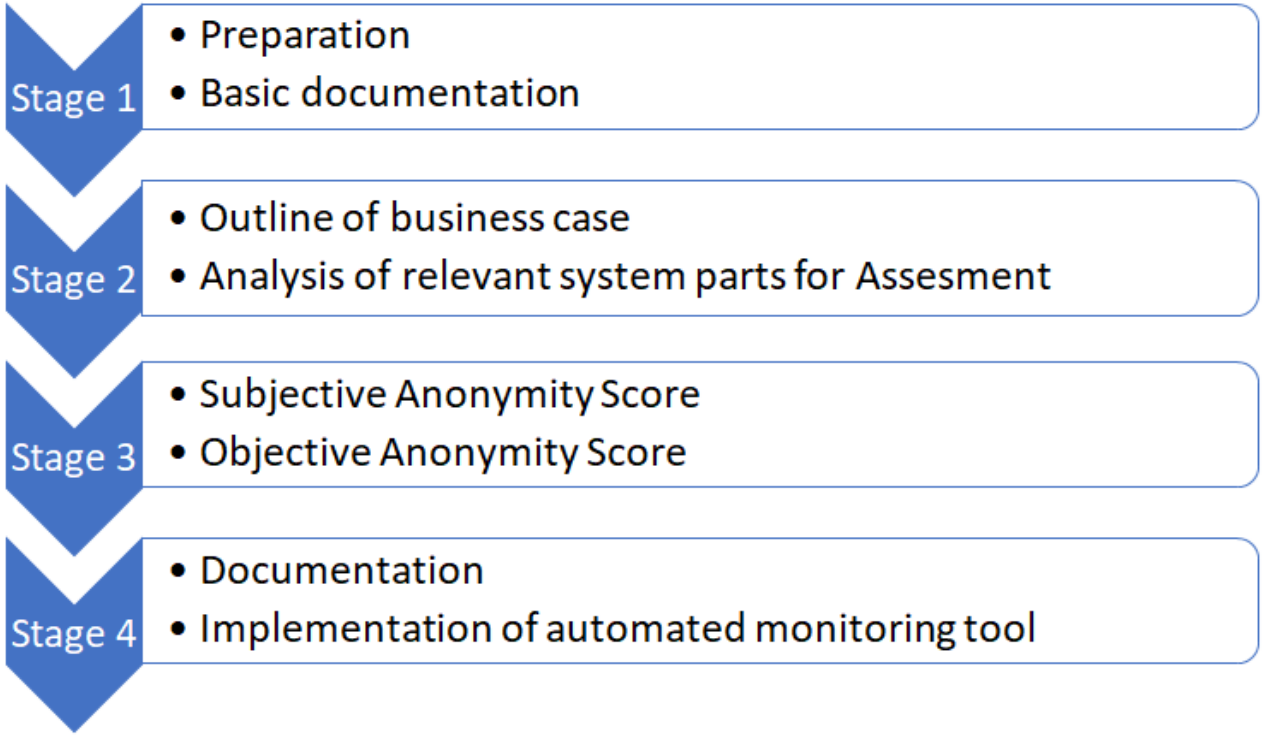
As soon as personal data is processed, European data protection law (esp. the GDPR) provides very strict rules that must be observed by data controllers and processors. This leads to a variety of problems, especially in the

Grundarchitektur Pflegerobotik

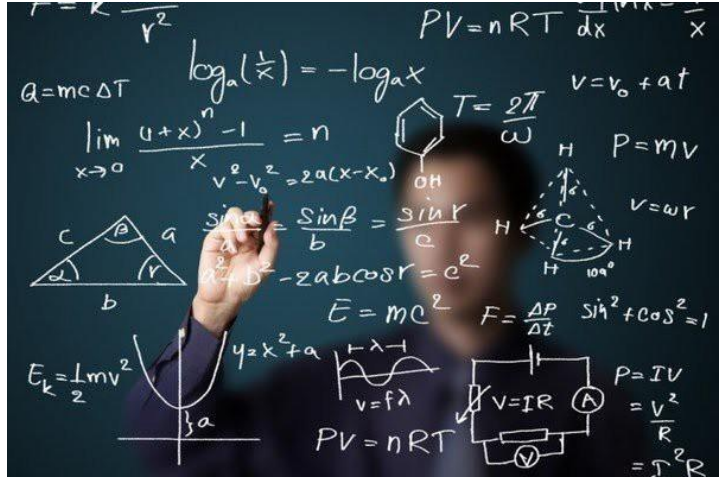


Methodische Herangehensweise

Ablauf eines Anonymity Assessment



Übersetzen der DSGVO in Gleichungen



Rechtlich relevante Überlegungen

- Pseudonyme Daten als Unterkategorie von personenbezogenen Daten - und Quantenzustand zwischen identifizierbaren und anonymisierten Daten
- Relativer Ansatz (Breyer-Case ECJ and Recital 26 GDPR)
- Kein allgemeiner Schwellenwert - DSGVO definiert keine klaren Kriterien zum klassifizieren von (strukturierten großen) Datensätzen als “personenbezogene Daten” oder "nicht-personenbezogene Daten”
- Kaum quantifizierbare Terminologie: “die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind”
- Mit rechtlichen Mitteln durchsetzbar (Breyer)

Binäre Bestimmung des Anwendungs-bereichs

Personenbezogenes Datum:
Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen;
als identifizierbar wird eine natürliche Person angesehen, die **direkt oder indirekt**, insbesondere mittels *Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen*, die Ausdruck der (...) Identität dieser natürlichen Person sind, identifiziert werden kann

anonyme Informationen = Informationen, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann.

Graduelle Spezifikation

alle Mittel (...), die von dem Verantwortlichen oder einer anderen Person **nach allgemeinem Ermessen wahrscheinlich genutzt werden**, um die natürliche Person direkt oder indirekt zu identifizieren, wie beispielsweise das Aussondern

alle **objektiven Faktoren**, wie die *Kosten* der Identifizierung und der dafür erforderliche *Zeitaufwand* + die zum Zeitpunkt der Verarbeitung *verfügbare Technologie und technologische Entwicklungen* berücksichtigen

Definitionen

Anonym: …aus ihrer Beschaffenheit heraus keinen Personenbezug enthalten…

Anonymisiert: …deren **Personenbezug entfernt** wurde, so dass i.d.R. eine Wiederherstellung des Personenbezugs möglich ist, aber mit unverhältnismäßig hohen Kosten und Zeitaufwand verbunden wäre…

Pseudo-anonymisiert: … die **stark pseudonymisiert** wurden, so dass eine Wiederherstellung des Personenbezugs möglich ist, aber die Verhältnismäßigkeit der Kosten und Zeitaufwand ungeklärt ist…

Ableitung von Anonymität aus der DSGVO

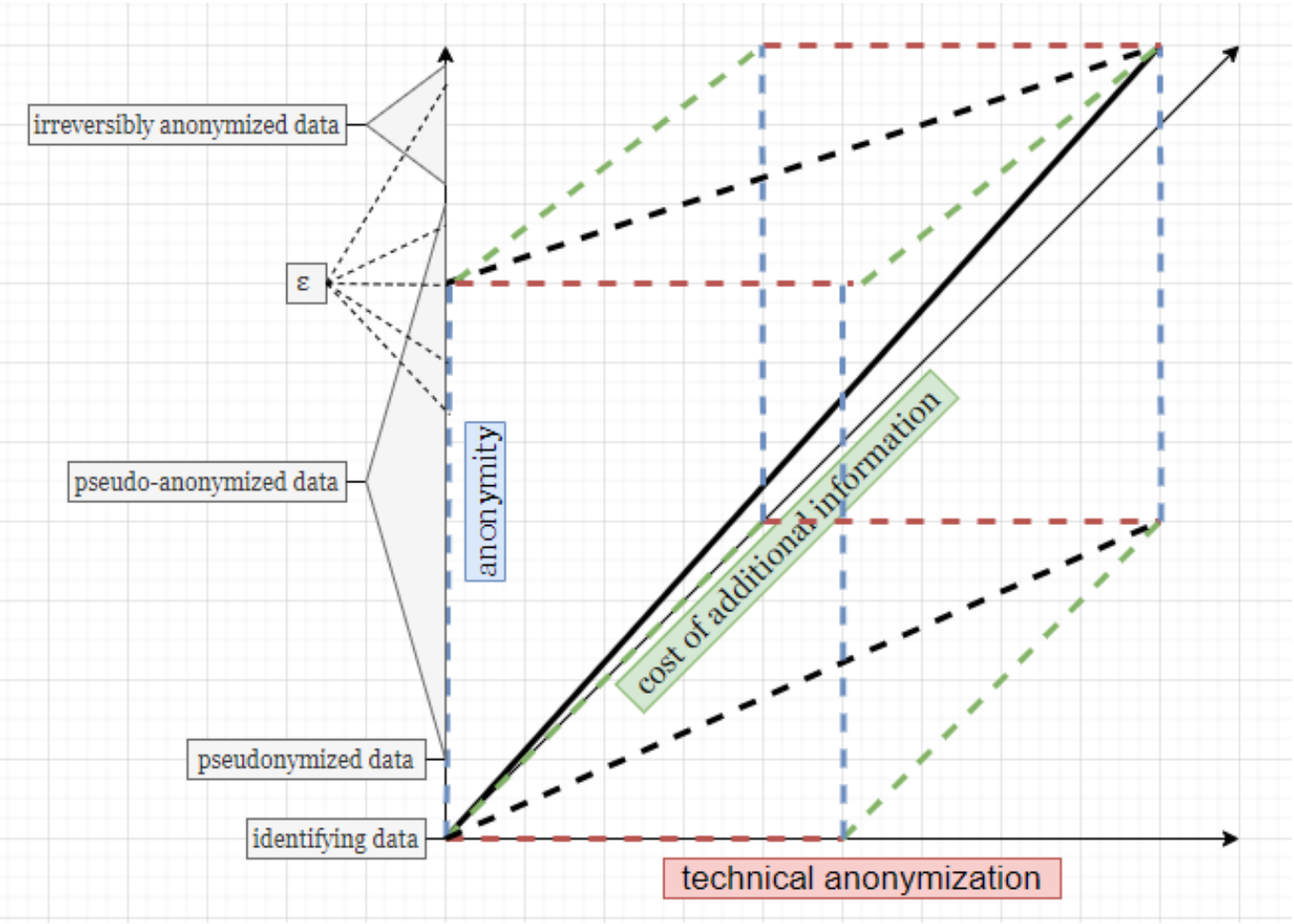
Grad der Pseudo-Anonymität = f(objektiver Teil, subjektiver Teil)

Formel für objektivierbaren Teil von Anonymität

$$(5)\ d_{ps\prime} = d_a \ \ if, and\ only\ if\ \exists\varepsilon:\ P(\ f(d_{ps\prime}, i)\ =\ d_p)\ <\ \varepsilon$$

Formel für subjektiven Teil von Anonymität

$$SAS\ = (c_i \times t_i)\ \div (c_a \times t_a)$$


irreversibly anonymized data
ε
pseudo-anonymized data
pseudonymized data
identifying data
anonymity
cost of additional information
technical anonymization

Differenzieren der Aspekte von Pseudo-Anonymität

Objective Anonymity Score (OAS)

bestimmt die **verbleibenden Risiken** der **(Re-)Identifizierung** einer natürlichen Person nach **objektiven** statistischen **Maßstäben**, insbesondere k-Anonymität, l-Diversität und t-Closeness.

Subjective Anonymity Score (SAS)

ist ein Indikator für die **relative Anonymität** der von einem für die Verarbeitung Verantwortlichen oder Auftragsverarbeiter durchgeführten Verarbeitung: Er berücksichtigt die **Kosten** und die **Zeit**, die für eine erfolgreiche Re-Identifizierung des fraglichen Datensatzes erforderlich sind, wobei die **verfügbaren Arbeitskräfte** und das **Kapital** des für die Verarbeitung Verantwortlichen / Verarbeiters berücksichtigt werden.

Stand der Technik

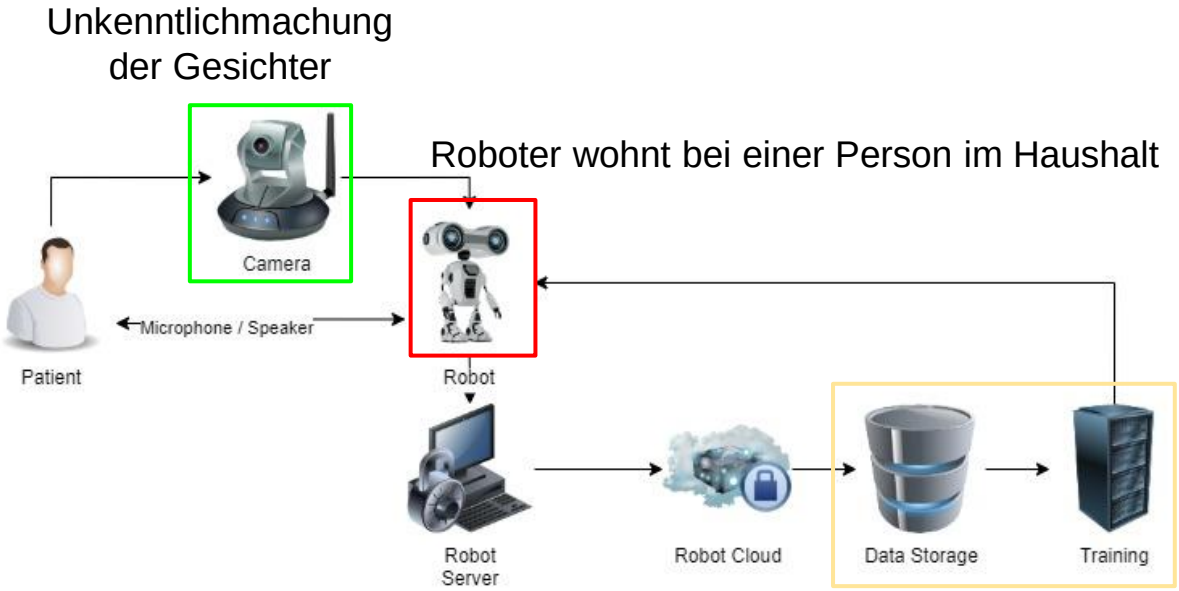
Offensichtliche Aspekte

- Anonymisierung ist Verarbeitung personenbezogener Daten.
- Anonymisierung muss unter Einhaltung der DSGVO erfolgen.
- Anonymisierung kann nur unter sorgfältiger und nachvollziehbarer Dokumentierung effektiv und anerkannt werden.
- Anonymität ist eine objektive momentane Eigenschaft von Daten.
- Das Zusammenführen mehrerer anonymer Datensätze kann einen Datensatz personenbezogener Daten erzeugen.
- Das Verarbeiten eines anonymen Datensatzes kann den Tatbestand einer Verarbeitung personenbezogener Daten erfüllen.

Aspekte der Anonymisierung

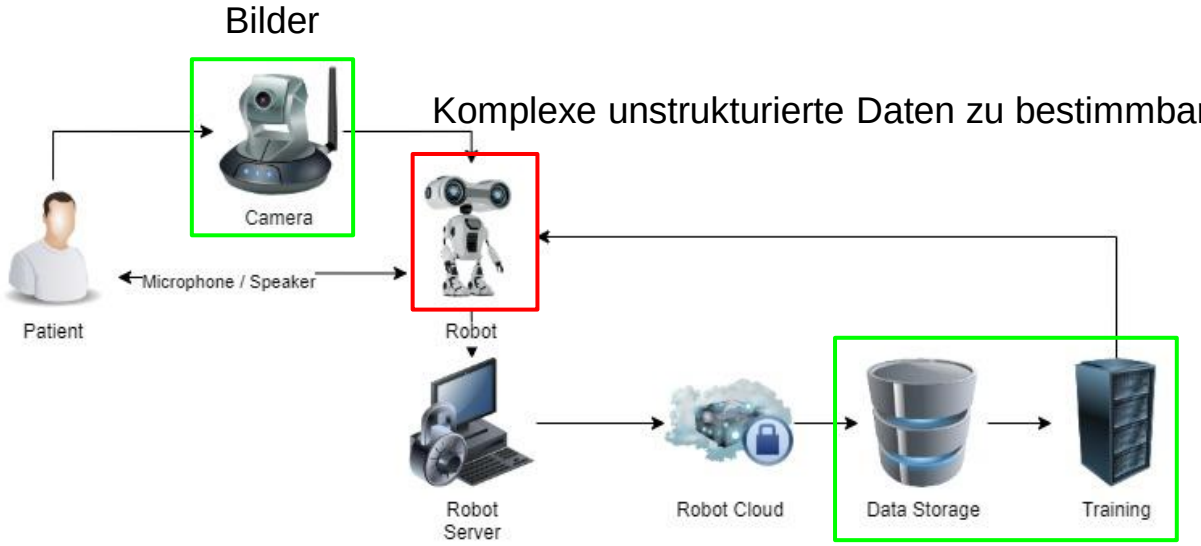
1. Realisierbarkeit der Anonymisierung
2. Struktur der Daten (z.B. Diffix)
3. Verarbeitungstechnologie
4. Risiko für Betroffene
5. Resultate der Verarbeitung (Zweck)

Beispiel: Realisierbarkeit



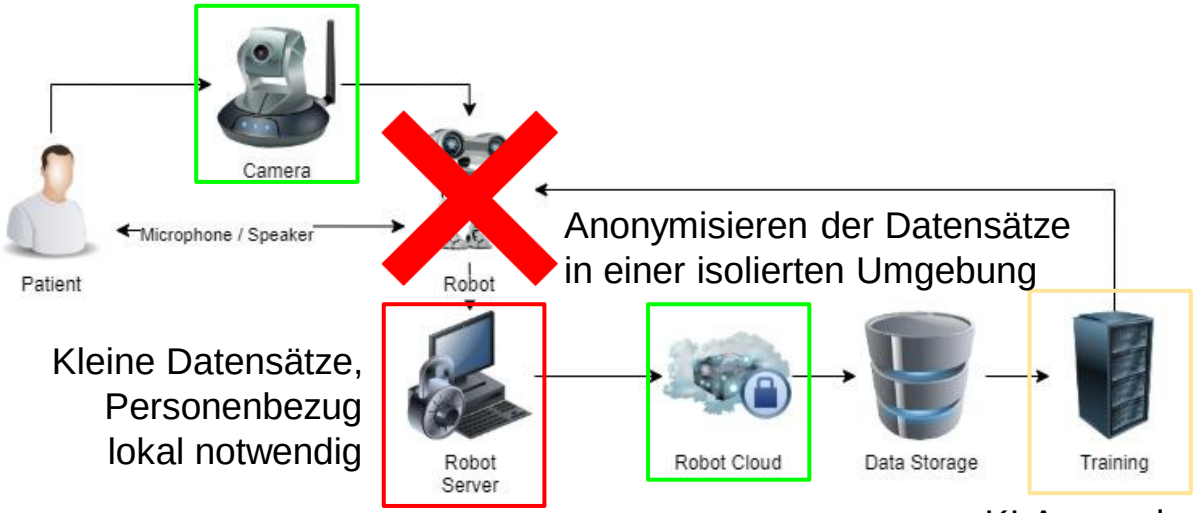
Grundsätlzlich bei Trainingsdaten möglich
Risikoanalyse notwendig

Beispiel: Struktur der Daten



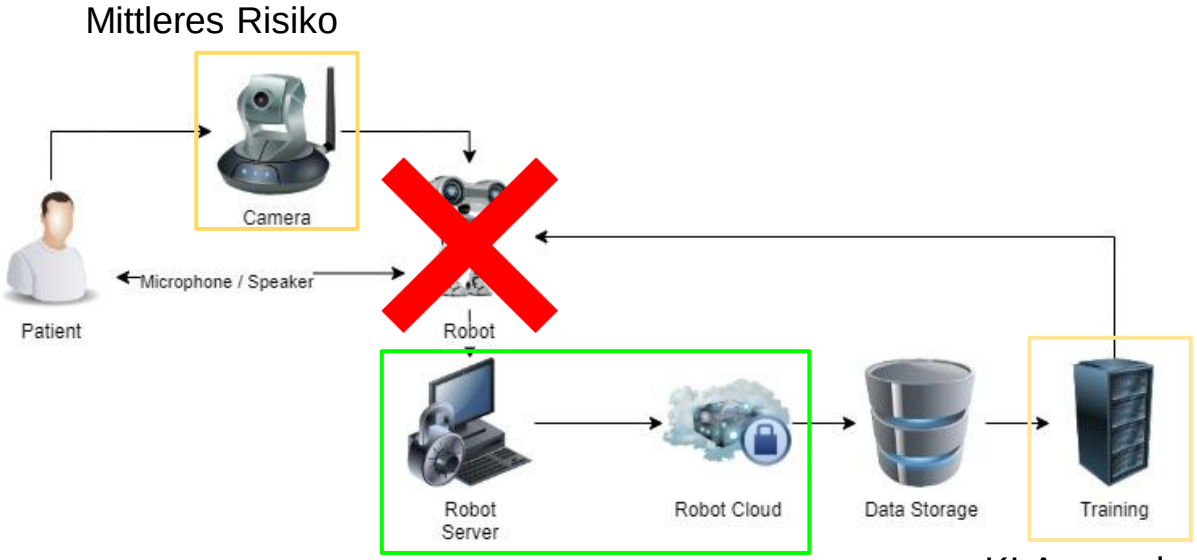
Beispiel: Verarbeitungstechnologie

Kamera, keine Verbindung zum Internet, Verpixelung auf dem Gerät



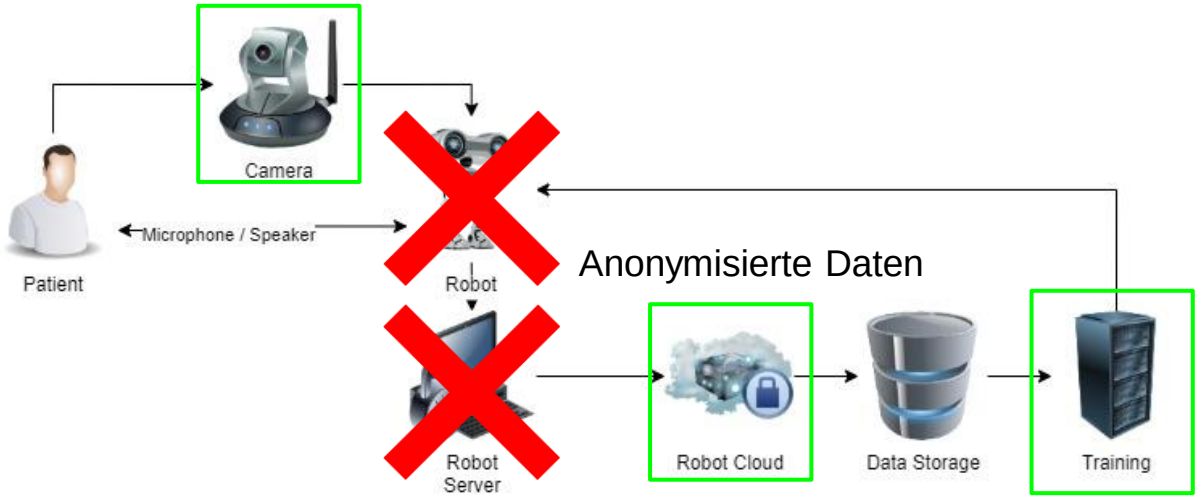
KI-Anwendung:
Grundsätzlich
hohes Risiko

Beispiel: Risiko für Betroffene



Beispiel 1: Resultate der Verarbeitung (Zweck)

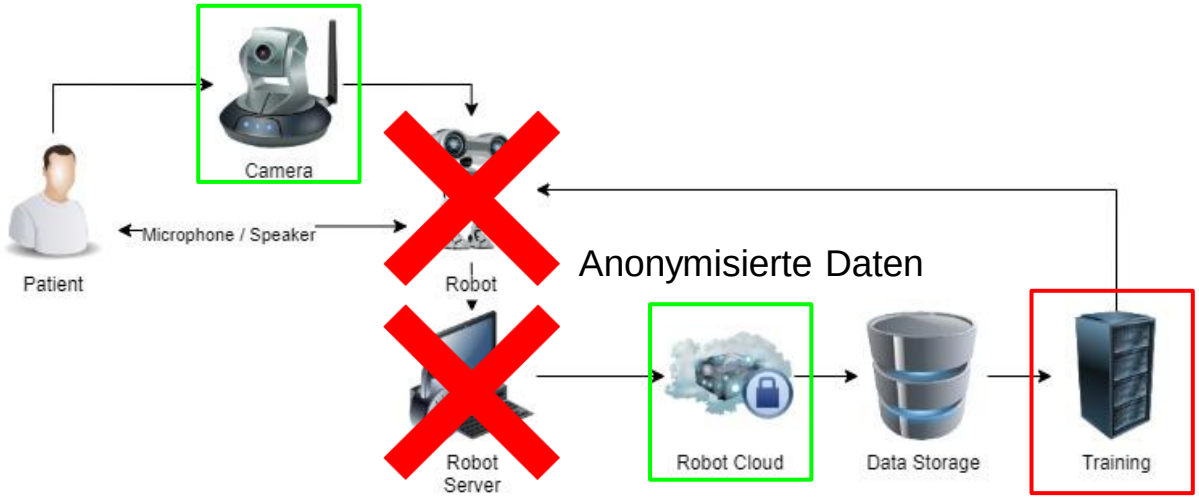
Verpixelte und ausgewertete Bilder zum Verbessern der Arbeitsabläufe des Roboters



Verbesserte Mustererkennung:
z.B. yoga vs. hinfallen

Beispiel 2: Resultate der Verarbeitung (Zweck)

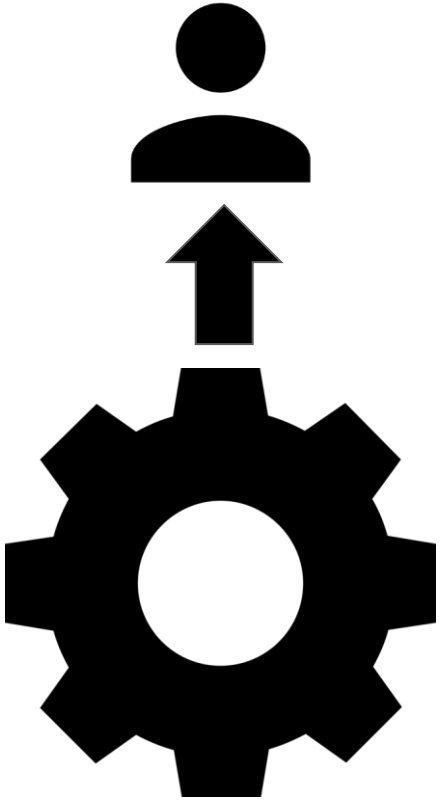
Verpixelte und ausgewertete Bilder zum Verbessern der Arbeitsabläufe des Roboters



Personalisierte Werbung,
keine Einwilligung vorhanden.

Anforderungen an Anonymisierung

Einzelfallanalyse



Anforderungen an Anonymisierung

Muss im **Einzelfall** geprüft und **dokumentiert** werden

Ausreichend hohes **K** als Mindestanforderung an **Anonymität**

Datensatz muss ausreichende Größe **N** haben

Werkzeuge zur Messung des **Risikos** der **Re-identifizierung** (z.B. open-diffix.org)

Alle Aspekte der Anonymisierung müssen berücksichtigt werden

Ergebnis der Verarbeitung, darf keine personenbezogenen Auswirkungen haben.

Die schwierigen Fragen

- Ab welchem k besteht ein angemessenes Schutzniveau für Betroffene?
- Welche Metriken sind am besten geeignet, um OAS zu messen?
- Wie geht man mit unzureichender Anonymisierung um?
- Sollte die Verarbeitung anonymisierter Daten meldepflichtig sein?
- Reicht eine Beurteilung einzig und alleine an der Beschaffenheit der Daten?
- Wie sieht es mit unstrukturierten, wachsenden Datensätzen aus?

Nächste Schritte

Ausgestaltung des Anonymity Assessments als **Gutachten**

Zertifizierungsmechanismus nach Art. 42 DSGVO für **Anonymisierung**

- ➔ Zertifierung der **Verarbeitungstätigkeit**: Anonymisierung
- ➔ Zertifizierung der Verarbeitung **anonymisierter** Daten bei **hohem Risiko**

Missbrauchspotentiale eindämmen

open diffix

RAILS.
Robotics & AI Law Society

DIN

TUM MSRM
Geriatronics