School of Management and Law
Institut für Unternehmensrecht

Zürcher Hochschule
für Angewandte Wissenschaften

zh
aw

PERSONENDATEN IM ZEITALTER VON MACHINE LEARNING – EINE ERSTE ANNÄHERUNG

20. Februar 2025, IRIS 2025 in Wien

Dr. iur. Philip Glass
Fachstelle für Datenschutz- und IT-Recht

Personendaten

Zürcher Hochschule
für Angewandte Wissenschaften

ART. 5 BST. A DSG

- Angaben über bestimmte oder bestimmbare Personen

⇒ 4 Informationskategorien: Angaben, Person, Verknüpfung, Identifizierung
  - Personendaten sind stets ein Datenmuster (das erhalten werden soll)
    - Kern des Musters ist die Verknüpfung mit Personen (Personenbezug)
    - Personendaten sind stets eine Form der Bearbeitung
      - zB auch data at rest ist Bearbeitung i.S. des Datenschutzrechts

    - Personenbezug: Zusammenspiel von syntaktischen und semantischen Informationen in Datenform
      - Passive und aktive Verknüpfung

Datenschutzrechtlicher (identifizierender) Personenbezug

Zürcher Hochschule für Angewandte Wissenschaften

zh aw

FORMEL PERSONENBEZUG IM CH-RECHT (BOTSCHAFT DSG 2017, BBL 2017 6941, 7019)

«Ist der Aufwand für die Bestimmung der betroffenen Personen derart gross, dass nach der allgemeinen Lebenserfahrung nicht damit gerechnet werden muss, dass ein Interessent diesen auf sich nehmen wird […], liegt keine Bestimmbarkeit vor.»
(Botschaft DSG 1988, BBl 1988 II 444)

Vielmehr muss die Gesamtheit der Mittel betrachtet werden, die vernünftigerweise eingesetzt werden können, um eine Person zu identifizieren. Ob der Einsatz dieser Mittel vernünftig ist, muss mit Blick auf die Umstände, etwa den zeitlichen und finanziellen Aufwand für die Identifizierung, beurteilt werden. Dabei sind die zum Zeitpunkt der Bearbeitung verfügbaren Technologien und deren Weiterentwicklung zu berücksichtigen.

⇒ Setzt eine gewisse Berechenbarkeit des Risikos voraus

Syntaktischer und semantischer Personenbezug

Zürcher Hochschule
für Angewandte Wissenschaften

zh
aw

ZWEI VERSCHIEDENE ASPEKTE VON INFORMATION

- Syntaktischer Personenbezug
  - Logische Zuweisung von Daten zu Personen mittels Syntax
  - «Verknüpfung» von Daten mit Personen mittels «Identifikatoren»
    - aktive / passive Verknüpfung
      - aktiv: Personendaten werden initial generiert
      - passiv: Nutzung bestehender Verknüpfungsmuster

- Semantischer Personenbezug
  - Bedeutungsgehalt der Daten bezieht sich notwendigerweise auf Personen (Menschen)

$\Rightarrow$ Betreffen beide nicht notwendigerweise (für jeden) bestimmbare Personen

Syntaktischer und Semantischer Personenbezug

IM EINZELNEN

- Syntaktische Komponente
  - Singularisierung durch syntaktisch korrekte Verknüpfung von Daten im Hinblick auf eine Person
    - benötigt Platzhalter für diese Person (beliebige Bezeichnung denkbar)
    - systeminterne Identifizierung

- Semantische Komponente
  - Semantische Übersetzung der syntaktischen Verknüpfung
    - erkennen der Bedeutung als Personendatum einer identifizierten Person)
    - Identfikatoren aus der Systemumwelt
      - Externe Identifizierung

Beobachtungen in Zusammenhang mit ML

Zürcher Hochschule für Angewandte Wissenschaften

zh aw

AUSGANGSLAGE

- ML: Mehrdimensionale stochastische Analysen transformieren Input zu Output
  - Output nicht notwendigerweise empirisch belastbar
  - mit statistischen Methoden berechnete Prognose generiert aus Input
  - syntaktische Verknüpfungen innerhalb des Modells unklar (noch?)
  - semantische Interpretation von Verknüpfungen innerhalb des Modells irrelevant (noch? AGI... :))

- Komplexe syntaktische Verknüpfungen; semantisch kaum interpretierbar (Opazität des Modells)
  - generiert jeweils «neue» (Personendaten)Daten (nicht: recall)
  - bestenfalls ex post nachvollziehbar
  - Unterscheidung aktive/passive Verknüpfung irrelevant
    - unklar, ob bearbeiten oder erheben von Personendaten

Beobachtungen in Zusammenhang mit ML

Zürcher Hochschule für Angewandte Wissenschaften

BEARBEITUNGSKONSTELLATIONEN

- Input Personendaten /Daten => Output Personendaten
  - Bearbeitung von Personendaten /Profiling (grundsätzlich rechtlich erfassbar)
- Auslesen von Trainingsdaten
  - umstritten ob durch recall oder generiert (grundsätzlich generiert)
  - Erfolg folgt offenbar nicht linear dem Aufwand.

⇒ Risikoprognose nach der klassischen Formel «fraglich».

- Lösung: ML-Modelle sollen als Pseudonyme für Trainingsdaten mit Personenbezug behandelt werden
  ⇒ Informationspflichten etc.