



# Machine Learning und die Transparenzanforderungen der DS-GVO

Leitfaden

[www.bitkom.org](http://www.bitkom.org)

**bitkom**

## Herausgeber

Bitkom  
Bundesverband Informationswirtschaft,  
Telekommunikation und neue Medien e. V.  
Albrechtstraße 10 | 10117 Berlin  
T 030 27576-0  
bitkom@bitkom.org  
www.bitkom.org

## Ansprechpartner

Rebekka Weiß, LL.M. | Bereichsleiterin Datenschutz & Verbraucherrecht  
T 030 27576-116 | r.weiss@bitkom.org

## Verantwortliche Bitkom-Gremien

AK Artificial Intelligence  
AK Datenschutz

## Autoren

Felix Bauer | Aircloak  
Stefan Buchberger | intelligent views  
Dr. Andreas Dewes | 7scientists  
Dr. Jörg Friedrichs | Deutsche Telekom  
Dr. Alexander Motzek | Lufthansa Industry Solutions  
Nicolas Sartor | Aircloak  
Dr. Lars Schwabe | Lufthansa Industry Solutions  
Thoralf Schwanitz | Google  
Rebekka Weiß, LL.M. | Bitkom

## Titelbild

© Fotolia.com – mypokcik

## Copyright

Bitkom 2018

Diese Publikation stellt eine allgemeine unverbindliche Information dar. Die Inhalte spiegeln die Auffassung im Bitkom zum Zeitpunkt der Veröffentlichung wider. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen. Eine Verwendung liegt daher in der eigenen Verantwortung des Lesers. Jegliche Haftung wird ausgeschlossen. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen beim Bitkom.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung und Executive Summary</b>	<b>3</b>
<b>2</b>	<b>Einführung in die Künstliche Intelligenz und das Machine Learning</b>	<b>5</b>
<b>3</b>	<b>Use Case: HR/Recruiting</b>	<b>15</b>
3.1	Problemdefinition	15
3.2	Risiko- und Nutzenbewertung sowie Zieldefinition	16
3.3	Auswahl geeigneter Daten zum Training des Systems	17
3.4	Untersuchung der Datenqualität und möglicher Probleme	18
3.5	Umsetzung eines Machine Learning Verfahrens	19
3.6	Evaluierung und kontinuierliche Überwachung der Ergebnisse	19
3.7	Datenschutzrechtliche Anforderungen	20
3.8	Weitere rechtliche Anforderungen und Methoden zur Umsetzung	20
<b>4</b>	<b>Machine Learning und die Datenschutzgrundverordnung</b>	<b>22</b>
4.1	Transparenzgrundsatz und Informationspflichten	22
4.2	Wie muss über Zwecke der Verarbeitung belehrt werden?	23
4.3	Artikel 6 Absatz 4 und noch unbekannte Nutzungszwecke	24
4.5	Rechtsgrundlagen	25
4.5.1	Datenverarbeitung für ML auf Basis der Rechtsgrundlage der Vertragserfüllung	25
4.5.2	Rechtliche Grundlagen	25
4.5.3	Leistungen, bei denen ML in vertraglich vereinbarter Art und Weise funktionieren muss	26
4.5.4	Fehler- und Mängelbeseitigung, um die Leistung auf dem vertraglich vereinbarten Niveau zu halten	27
4.5.5	Vertraglich vereinbarte kontinuierliche Verbesserung	27
4.6	Gefahr der Restriktion durch die ePrivacy-Verordnung	28
<b>5</b>	<b>Privacy Enhancing Technologies</b>	<b>30</b>
5.1	Pseudonymisierung	31
5.2	K-Anonymisierung	33
5.2.1	Kritik	35
5.2.2	K-Anonymisierung und Machine Learning	35
5.3	Differential Privacy	36
5.3.1	Differential Privacy ist kein Algorithmus	38
5.3.2	No Free Lunch	39
5.3.3	Zusammenfassung	41

# Abbildungsverzeichnis

Abbildung 1: KI-Landschaft, adaptiert von intelligent views GmbH _____	7
Abbildung 2: Symbolische / Wissensbasierte KI, adaptiert von intelligent views GmbH _____	8
Abbildung 3: Subsymbolische / Datenbasierte KI, adaptiert von intelligent views GmbH _____	10
Abbildung 4: Layerstruktur eines neuronalen Netzes, adaptiert von intelligent views GmbH _____	11
Abbildung 5: Komplexere Layerstruktur des Deep Learning, adaptiert von intelligent views GmbH _____	12
Abbildung 6: Hybride KI: Verknüpfung zwischen statistischen Lernmethoden und großen Wissensrepräsentationen (subsymbolischer KI), adaptiert von intelligent views GmbH _____	14
Abbildung 7: Entscheidungsbaum für einen Prozess der von zwei Gruppen von Personen, adaptiert von 7scientists _____	21
Abbildung 8: Beispiele Pseudonymisierung, adaptiert von Aircloak _____	31
Abbildung 9: Beispieltabellen K-Anonymisierung, adaptiert von Aircloak _____	34
Abbildung 10: Datenfluss mit und ohne Differential Privacy Ansätze, adaptiert von Lufthansa Industry Solutions _____	37

# Verzeichnis der Tabellen

Tabelle 1: Vor- und Nachteile von Pseudonymisierung _____	32
Tabelle 2: Vor- und Nachteile der K-Anonymität _____	35

# 1 Einleitung und Executive Summary

Diese Publikation adressiert Entscheidungsträger in der Politik und in den Datenschutzbehörden, private Verbraucher sowie die breite Öffentlichkeit und die Medien. Sie wendet sich ebenfalls an die Nutzer von Machine Learning (ML) in der Wirtschaft, deren Anbieter sowie Datenschutzbeauftragte.

Sie verfolgt das Ziel, Leitlinien für den datenschutzkonformen, daten- und verantwortungsbewussten Einsatz von Machine Learning zu formulieren und eine rechtssichere Handhabung aufzuzeigen. Die Publikation soll auch zu mehr Transparenz im Bereich Machine Learning beitragen, denn Information und Transparenz schaffen die Basis für Vertrauen, das in Zeiten der immer technischer werdenden Datenverarbeitungsprozesse und der automatisierten Entscheidungen unerlässlich ist.

Da Machine Learning in absehbarer Zeit die Art und Weise revolutionieren wird, wie wir arbeiten, lernen, kommunizieren, konsumieren und leben, verdient die Technologie eine besondere Betrachtung. ML-Anwendungen können soziale Inklusion voranbringen, Sprachdefizite überbrücken und Menschen mit eingeschränkter Mobilität helfen, ihren Alltag selbstbestimmter zu gestalten. Verbesserte medizinische Diagnostik, effizientere Energieversorgung oder auch Verkehrsplanung und Parkplatzsuche, Bild- und Spracherkennung sind nur einige Beispiele der großen und auch bereits angewandten Möglichkeiten von ML. Auch um den Eintritt von Risiken auf die Produktionsversorgung, auf die Börsenkursentwicklung oder die Volkswirtschaft frühstmöglich zu antizipieren eignet sich ML – in diesen Fällen wird beispielsweise die Analyse von Risikoindikatoren ermöglicht, deren Zahl und mitunter sehr indirekten Wirkzusammenhänge für den menschlichen Verstand und bisherige Verfahren zu komplex sind.

Machine Learning wirft aber auch besondere Fragen auf, wenn es um die Verarbeitung personenbezogener Daten geht oder wenn Kausalzusammenhänge nicht korrekt interpretiert werden. Vertieft beschäftigt sich derzeit mit diesen Fragestellungen die Bitkom Publikation zum verantwortungsvollen Einsatz von automatisierten Entscheidungen und KI. Der vorliegende Leitfaden soll vor allem die datenschutzrechtlichen Fragestellungen adressieren, die sich mit den Anforderungen an Transparenz und den Rechtsgrundlagen der DS-GVO befassen, auf die die Anwendungen gestützt werden können. Der Leitfaden soll damit einen Beitrag zum Diskurs liefern und die Vereinbarkeit von innovativen Technologien mit der DS-GVO verdeutlichen, jedoch nicht ohne auch auf die Hemmnisse hinzuweisen, die durch einen zu restriktiven Rahmen entstehen.

Das Papier konzentriert sich dabei auf vier Schwerpunkte:

- Die Einordnung des Machine Learnings in das System aus Algorithmen, automatisierten Entscheidungen und KI. Es werden technische Verfahren erläutert und so Transparenz in das Verfahren und seine Anwendungen gebracht. Verständnis der Technologie ist sowohl für die Rechtsanwendung als auch für weitere Regulierungsbestrebungen des Datenschutzrechts aber auch z. B. im Bereich Text-and-Data Mining unerlässlich.
- Anhand eines Use Cases wird der Nutzen für die Verbraucher und die Gesellschaft insgesamt gezeigt, den Machine Learning mit sich bringt.
- Den dritten Schwerpunkt bilden datenschutzrechtliche Fragestellungen, die im Zusammenhang mit Machine Learning diskutiert werden müssen. Die DS-GVO gibt hier einen Rahmen vor ohne jedoch konkrete Anwendungen zu erlauben oder zu verbieten. Die Rechtsgrundlagen müssen daher sauber angewendet und voneinander abgegrenzt werden. Ein Einfügen der Technik in das System des Datenschutzes ist möglich und wird hier aufgezeigt.
- Im vierten Teil widmet sich das Papier einigen ausgewählten Methoden der Pseudonymisierung und Anonymisierung und zeigt damit Wege auf, wie Schutzmaßnahmen im Rahmen der DS-GVO und der technischen und organisatorischen Maßnahmen funktionieren können bzw. der Personenbezug sogar gänzlich gekappt werden kann und damit der Anwendungsbereich der DS-GVO verlassen wird.

## 2 Einführung in die Künstliche Intelligenz und das Machine Learning

Künstliche Intelligenz hat sich in letzten Jahren zum absoluten Topthema entwickelt. Die Anzahl an Veröffentlichungen, Marktstudien und die breit geführte Diskussion unterstreichen die Bedeutung und Aktualität. Vor allem Machine Learning Verfahren, die ein Teilgebiet der künstlichen Intelligenz wie Neuronale Netze oder Deep Learning sind, erleben aktuell einen Hype und werden gerne in einem Atemzug genannt.

Das richtige Maß an Transparenz, verantwortungsvollem Umgang und Nachvollziehbarkeit, sowie Klarheit über die Technologien und ein gemeinsames Verständnis sind die Wegbereiter für Akzeptanz und erfolgreichen Einsatz.

Mit ihrem am 18.07.2018 veröffentlichten Eckpunktepapier unterstreicht die Bundesregierung ihr Ziel, Deutschland als führenden Standort für die Forschung und Anwendung künstlicher Intelligenz ausbauen zu wollen:

»Die Bundesregierung sieht sich in der Pflicht, eine verantwortungsvolle und gemeinwohlorientierte Nutzung von KI in Zusammenarbeit mit Wissenschaft, Wirtschaft, Staat und der Zivilgesellschaft voranzubringen.«<sup>1</sup>

Dort heißt es auch: »Wir wollen Wertschöpfung aus der Anwendung von KI erzeugen, den Nutzen von KI für die Bürgerinnen und Bürger in den Fokus unserer Bemühungen stellen – sowohl auf der persönlichen, individuellen Ebene als auch auf der gesellschaftlichen – und insbesondere veränderungsbedingte Risiken minimieren, Systeme überprüfbar machen und unzulässige Diskriminierungen unterbinden.«

Auch wenn aktuell in aller Munde, ist KI nicht neu und hat bereits einige Höhen und Tiefen hinter sich. So gehen beispielsweise erste Arbeiten zu neuronalen Netzen der beiden Elektroingenieure Warren McCulloch und Walter Pitts bereits auf das Jahr 1943 zurück.<sup>2</sup>

1 [https://www.bmbf.de/files/180718%20Eckpunkte\\_KI-Strategie%20final%20Layout.pdf](https://www.bmbf.de/files/180718%20Eckpunkte_KI-Strategie%20final%20Layout.pdf)

2 <https://de.wikipedia.org/wiki/McCulloch-Pitts-Zelle>

Beschäftigt man sich mit künstlicher Intelligenz, so ist es sinnvoll, Ansätze, Techniken, Verfahren und Begrifflichkeiten zu differenzieren, einzuordnen und zu beschreiben, denn gerade in der aktuellen Diskussion mangelt es oft an Transparenz und begrifflicher Klarheit.<sup>3</sup>

#### Was ist KI?

Künstliche Intelligenz beschreibt Informatik-Anwendungen, deren Ziel es ist, intelligentes Verhalten zu zeigen. Dazu sind in unterschiedlichen Anteilen bestimmte Kernfähigkeiten notwendig: Wahrnehmen, Verstehen, Handeln und Lernen. Diese vier Kernfähigkeiten stellen die größtmögliche Vereinfachung eines Modells zur modernen KI dar: Wahrnehmen – Verstehen – Handeln erweitern das Grundprinzip aller EDV Systeme: Eingabe – Verarbeitung – Ausgabe.<sup>4</sup>

Man unterscheidet zwischen zwei Arten der künstlichen Intelligenz: die starke und die schwache. Ziel der starken KI ist es, eine Intelligenz zu erschaffen, die der des Menschen ebenbürtig ist oder diese sogar übertrifft. Bislang war dieses Unterfangen erfolglos und es ist fraglich, ob es je gelingen kann.

Spricht man heute von KI ist meist die schwache, anwendungsorientierte KI gemeint, bei der es darum geht, den Menschen intelligent beim Erreichen seiner Ziele zu unterstützen. Sie lässt sich am besten definieren, als Vermögen einzelne kognitive Aufgaben so gut wie oder gar besser als der Mensch zu bewältigen. Als Beispiele lassen sich Bild-, Text-, Sprach- und Muster-Erkennungen nennen, die bereits einen festen Platz im heutigen Alltag gefunden haben.

Betrachtet man die Verfahren der schwachen KI genauer, lassen sich diese in die symbolische (wissensbasierte) und statistische bzw. subsymbolische (datenbasierte) KI kategorisieren.

<sup>3</sup> Dem interessierten Leser sei an dieser Stelle eine weitere Bitkom Publikation empfohlen: das Periodensystem der künstlichen Intelligenz, welche als Navigationshilfe durch die Elemente und Einsatzgebiete künstlicher Intelligenz nützlich ist.

<sup>4</sup> Die Definition der Künstlichen Intelligenz ist dem Bitkom & DFKI Positionspapier: Künstliche Intelligenz – Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung, Seite 29 entnommen.



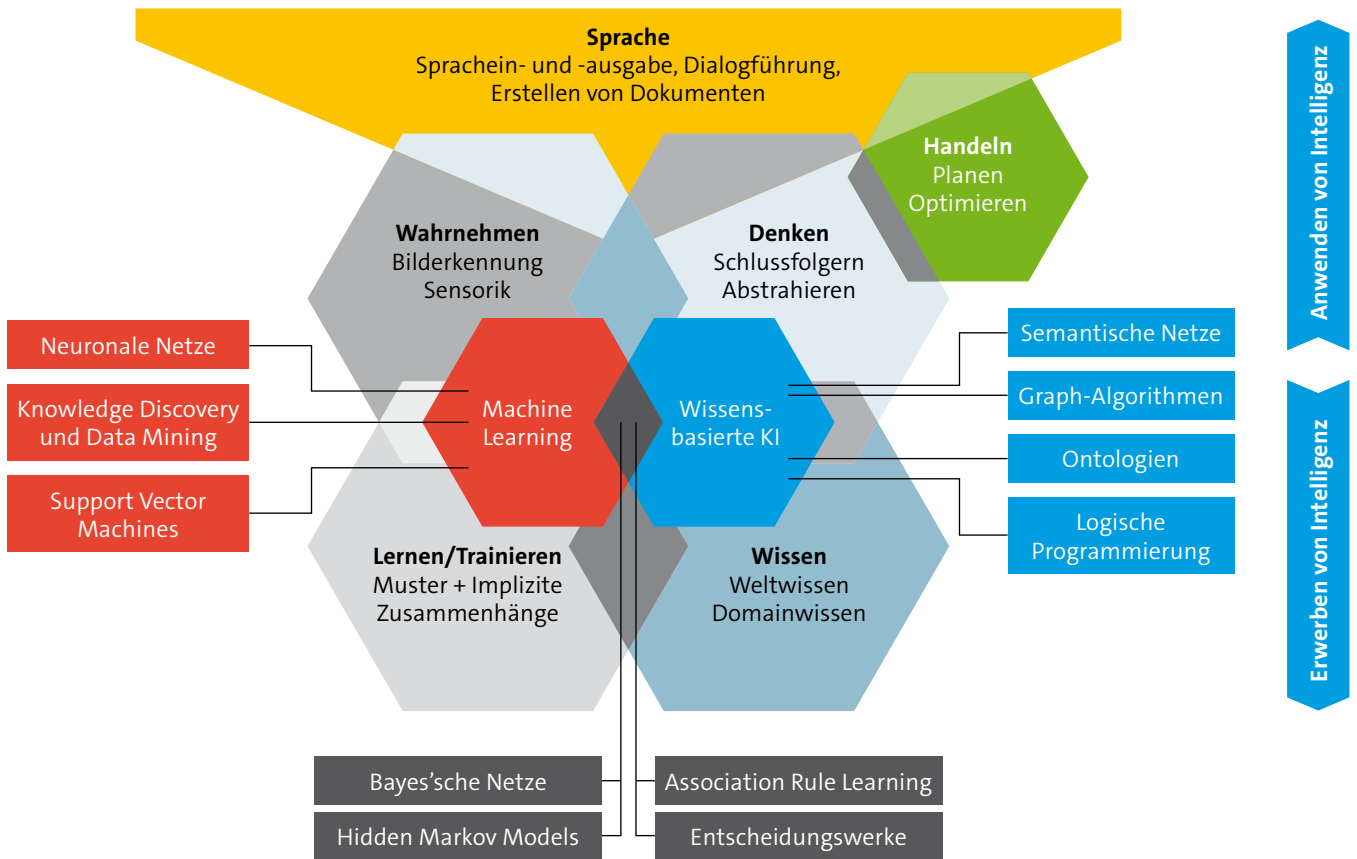


Abbildung 1: KI-Landschaft, adaptiert von intelligent views GmbH

Symbolisch bedeutet hier, dass Fakten, Ereignisse oder Aktionen mit konkreten und eindeutigen Repräsentationen erfasst werden. Wissen ist hier ein explizites Modell von Fakten und Zusammenhängen, das die Grundlage bewusster Intelligenzleistungen/Denkprozesse und Handlungen bietet. Der Mensch kann Wissen über Erkenntnis erwerben oder es kann ihm explizit beigebracht werden.

→ In der künstlichen Intelligenz wird das Wissen aktuell immer beigebracht.

Die symbolische KI schafft Modelle, mit denen Menschen und Systeme gleichermaßen souverän umgehen können und nähert sich den Intelligenzleistungen von einer begrifflichen Ebene her.

Als Verfahren der symbolischen KI sind hier zu nennen: Semantische Netze, Graph-Algorithmen, Ontologien oder die logische Programmierung.

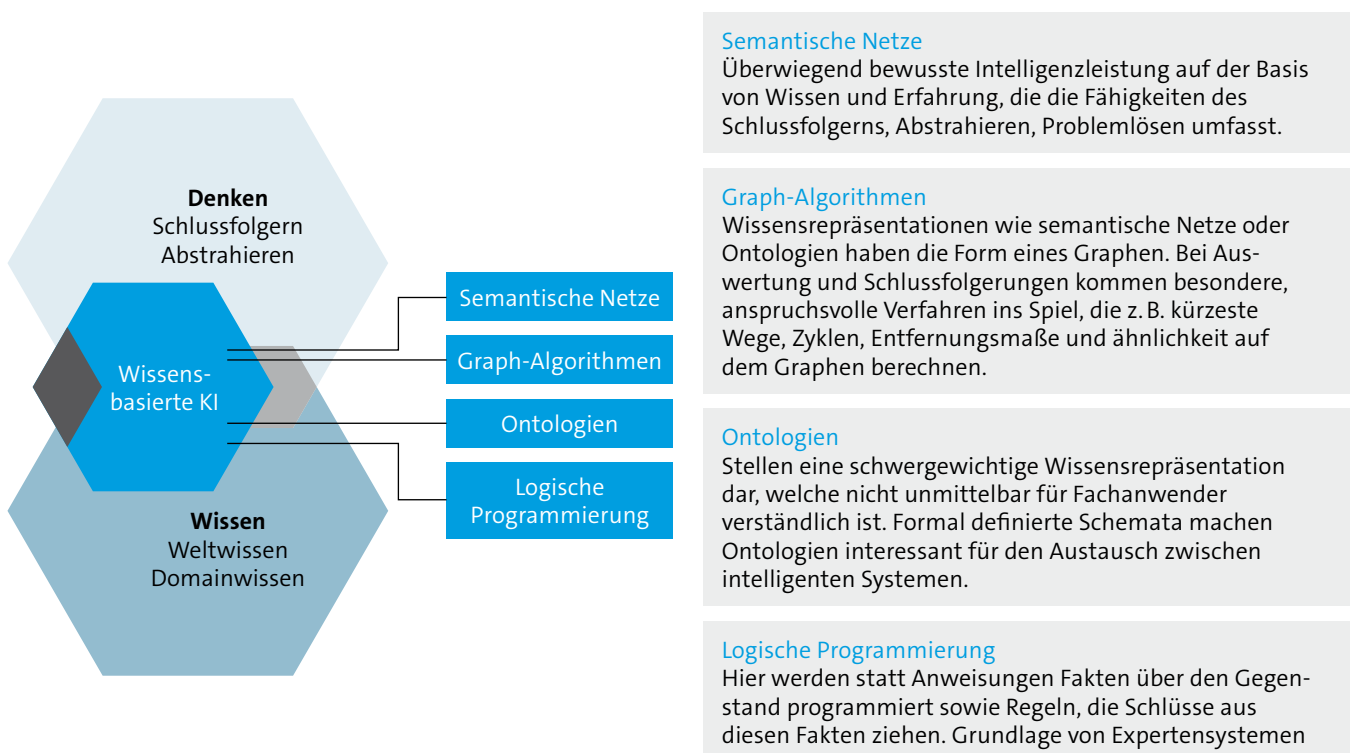


Abbildung 2: Symbolische/Wissensbasierte KI, adaptiert von intelligent views GmbH

Der Ansatz der statistischen KI erreicht die Intelligenzleistungen mit »Datenfütterungen«. Ein berechenbares Verhalten wird gelernt bzw. »antrainiert«, bietet allerdings keinen Einblick in die erlernten Lösungswege. Das Wissen ist hier implizit repräsentiert und gibt der subsymbolischen KI den Charakter einer Black Box. Vergleichbar ist es mit der Verarbeitung von Reizen bei Menschen.

Diese, daher auch als »Machine Learning« bezeichnete Kategorie, vereint Methoden der mathematischen Theorie und Optimierung, der Statistik und des Data Mining. Für den Trainingsvorgang und Lernerfolg sind größere Datenmengen sowie die geeignete Auswahl wichtig.

Das Lernen lässt sich in folgende Arten unterteilen:

- überwachtes Lernen (supervised learning)
- unüberwachtes Lernen (unsupervised learning)
- teilüberwachtes Lernen (semi-supervised learning)
- bestärkendes Lernen (reinforcement learning)
- aktives Lernen (active learning)

Bei der Methode des überwachten Lernens müssen die richtigen Antworten zu den Beispielen als sogenannte Labels mitgeliefert werden. Es wird also eine Funktion aus gegebenen Paaren von Ein- und Ausgaben trainiert. Ziel beim überwachten Lernen ist es, dass nach mehreren Rechengängen mit unterschiedlichen Ein- und Ausgaben die Fähigkeit antrainiert wird, Assoziationen herzustellen.

Hingegen reichen beim unüberwachten Lernen die rohen Beispieldaten aus, um grundlegende Muster in den Daten zu entdecken.

→ **Das teilüberwachte Lernen stellt eine Mischung aus überwachtem und unüberwachtem Lernen dar.**

Beim bestärkenden Lernen nutzen Maschinen Feedback aus ihrer Interaktion mit der Umwelt, um ihre zukünftigen Aktionen zu verbessern und Fehler zu verringern. Sie lernen durch Belohnung und Bestrafung eine Taktik, wie in potenziell auftretenden Situationen zu handeln ist, um den Nutzen des Systems – zu dem die Lernkomponente gehört – zu maximieren.

Aktives Lernen schließlich bietet dem System die Möglichkeit, für bestimmte Eingangsdaten die gewünschten Ergebnisse zu erfragen. Um die Anzahl von Fragen zu minimieren, erfolgt zuvor eine Auswahl relevanter Fragen mit hoher Ergebnisrelevanz durch das System selbst.

Mittlerweile gibt es eine Vielzahl von Trainingsverfahren bzw. konkrete Machine Learning Verfahren, die jeweils für unterschiedliche Aufgaben besonders gut geeignet sind.

### Neuronale Netze

Schwergewichtiges und meist mehrschichtiges, maschinelles Lernverfahren in Form von verbundenen Neuronen, nicht für Menschen direkt nachvollziehbar, daher Black-Box. Anhand einer großen Menge von Trainingsdaten werden für Eingaben bestimmte vorgegebene Ausgabewerte erlernt.

### Knowledge Discovery und Data Mining

Familie aller Verfahren, die in meist großen Datenmengen nach neuen Mustern und Regeln suchen. Viele Verfahren des maschinellen Lernens können auch zum Data Mining verwendet werden und umgekehrt.

### Support Vector Machines

Rein mathematisches und rechenintensives Verfahren zur Mustererkennung, bei welchem bekannte Objekte in einem Vektorraum repräsentiert werden, um passende Trenn-Ebenen zwischen verschiedenen Objekt-Kategorien zu berechnen.

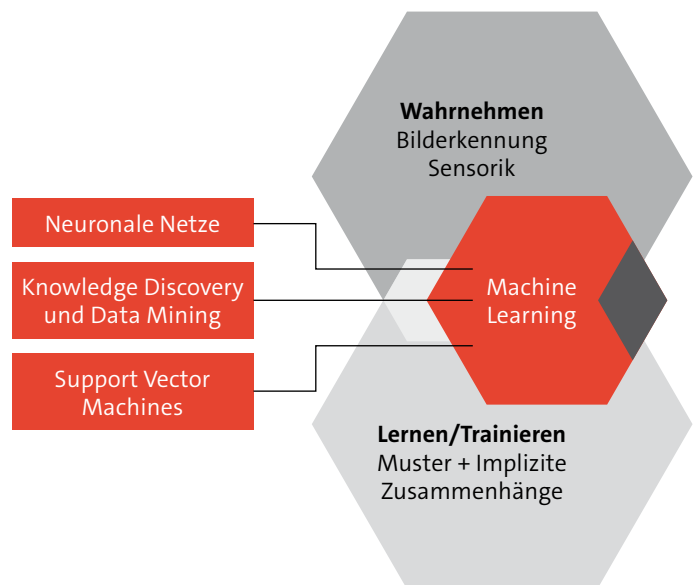


Abbildung 3: Subsymbolische / Datenbasierte KI, adaptiert von intelligent views GmbH

Auf die prominenten Vertreter des Machine Learning, nämlich die neuronalen Netze und Deep Learning, die eine spezielle Form des neuronalen Netzes sind, sei noch etwas tiefer eingegangen.

In Anlehnung an die grundlegenden Mechanismen des menschlichen Gehirns ist das Künstliche Neuronale Netz ein Verbund sehr einfacher Verarbeitungseinheiten, die Neuronen genannt werden. Diese Neuronen – oft auch als auch als Units, Einheiten oder Knoten bezeichnet – sind durch einseitige Kommunikationskanäle miteinander verbunden, über die sie Informationen aus der Umwelt oder von anderen Neuronen aufnehmen und an andere Units oder die Umwelt in modifizierter Form weiterleiten. Der Output eines Neurons kann zum Input des nächsten Neurons werden.

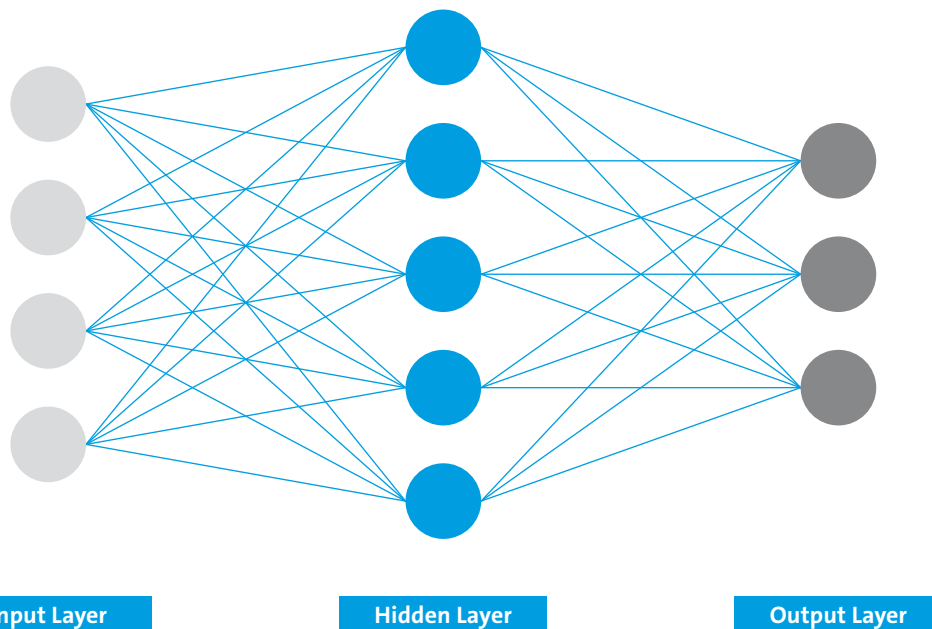


Abbildung 4: Layerstruktur eines neuronalen Netzes, adaptiert von intelligent views GmbH

»Übereinander« angeordnete Knoten/Neuronen fasst man als Schicht bzw. Layer zusammen.

Grundsätzlich wird zwischen Input-Layer-, Hidden-Layer-Neuronen und Output-Neuronen unterschieden. Der Input-Layer nimmt Informationen in Form von Mustern oder Signalen von der Außenwelt auf und der Output-Layer gibt Informationen und Signale als Ergebnis ab. Dazwischen befindet sich der Layer der Hidden-Neuronen, die interne Informationsmuster abbilden.

Die Stärke der Verbindung zwischen zwei Neuronen wird durch ein Gewicht ausgedrückt. Je größer das Gewicht, desto größer ist der Einfluss einer Unit auf eine andere Unit.

- Ein **positives Gewicht** bringt zum Ausdruck, dass ein Neuron auf ein anderes Neuron einen erregenden, stärkenden Einfluss ausübt.
- Ein **negatives Gewicht** bedeutet, dass der Einfluss hemmend ist.
- Ein **Gewicht von Null** besagt, dass ein Neuron auf ein anderes Neuron derzeit keinen Einfluss ausübt.

Die künstliche »Intelligenz« eines neuronalen Netzes ist also in den Verbindungen und deren Gewichtungen gespeichert. Während des Trainings des Neuronalen Netzwerks, verändern sich die Gewichtungen der Verbindungen, abhängig von den angewandten Lernregeln und erzielten Ergebnissen.

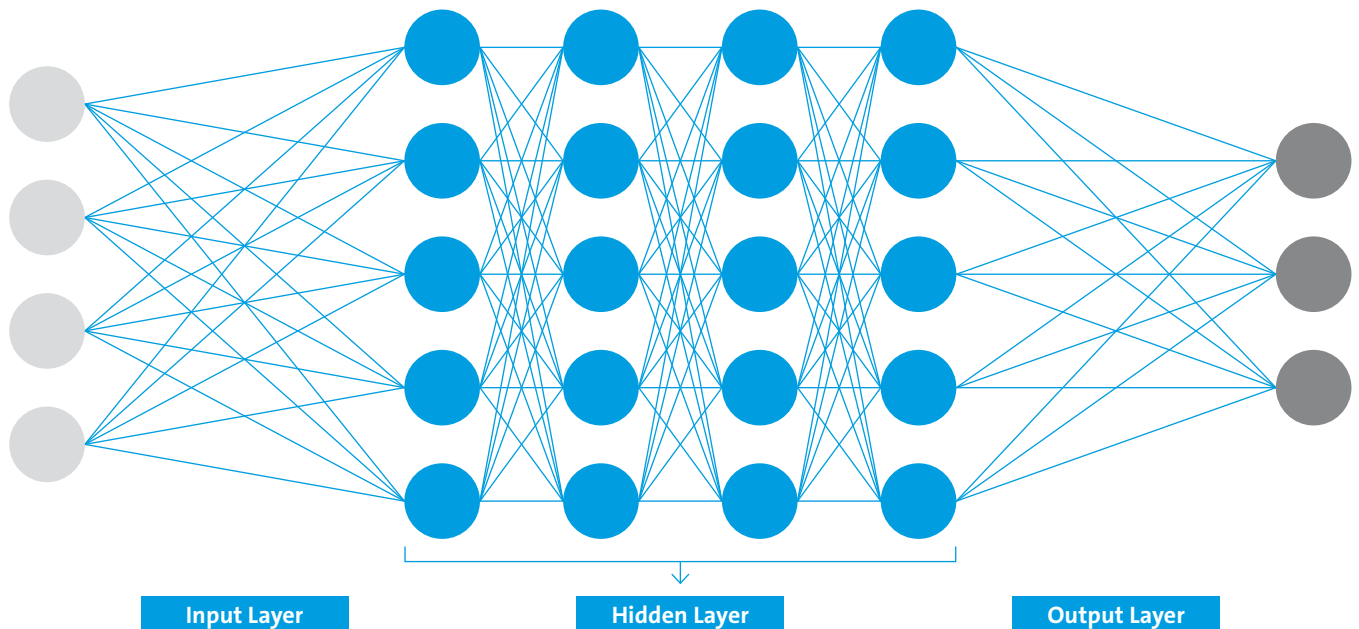


Abbildung 5: Komplexere Layerstruktur des Deep Learning, adaptiert von intelligent views GmbH

Eine optimierte, komplexere und rechenintensivere Form sind Netze mit Deep-Learning Topologien. Diese bestehen aus vielen Schichten von Neuronen, die typischerweise jeweils abstrahieren bzw. auf bestimmte Features, also Eigenschaften des Eingangssignals, ansprechen. Diese tiefe Schachtelung ist höchst leistungsfähig und für die Namensgebung »Deep Learning« verantwortlich.

In den letzten 10 Jahren verzeichnete das Lernen mit tiefen künstlichen neuronalen Netzen enorme Fortschritte, insbesondere in der Analyse von Bild- und Video-, Sprach- und Textdaten. Inzwischen können Systeme in einigen Fällen Gesichter und Objekte mit einer geringeren Fehlerquote identifizieren als Menschen und sogar Fachleute.<sup>5</sup>

Allerdings sind hier große Datenmengen und eine hohe Rechenleistung erforderlich, denn die Leistungsfähigkeit eines Neuronalen Netzes skaliert mit den Trainingsdaten. Deep-Learning lässt sich also nur mit signifikant großen, erschlossenen Datenmengen sinnvoll einsetzen.

5 <https://arxiv.org/pdf/1502.01852v1.pdf> Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification By Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun; Microsoft Research.

Machine Learning Verfahren im Allgemeinen und Neuronale Netze bzw. Deep-Learning im Besonderen erscheinen zum Teil wie eine Black Box, der es an Transparenz, Erklärbarkeit und Nachvollziehbarkeit der Ergebnisse mangelt.<sup>6</sup>

Genauso wenig wie wir bis heute die Vorgänge im menschlichen Gehirn verstehen, kann man beim Deep Learning sagen, warum eine bestimmte Entscheidung getroffen wird. Wissenschaftliche Studien haben insbesondere gezeigt, dass sich Netze für die Bilderkennung mit speziell präparierten Bildern teilweise leicht täuschen lassen.<sup>7</sup>

Legt man z. B. bestimmte Muster oder Rauschen über ein Bild, können Verkehrszeichen plötzlich nicht mehr erkannt werden. Desweiteren bergen die Trainingsdaten ein gewisses Risiko. Werden sie falsch gewählt oder manipuliert, kann das neuronale Netz schnell in eine negative Richtung trainiert werden.

Zukünftig werden KI-Verfahren bedeutsam sein, die eine Verknüpfung zwischen statistischen Lernmethoden und großen Wissensrepräsentationen (subsymbolischer KI) herstellen und Nachvollziehbarkeit, Verständlichkeit und Erklärbarkeit erlauben.

Die Kombination beider Welten hat das Ziel, die sogenannte »semantische Lücke« zu schließen. Diese entsteht dort, wo intuitives Weltwissen und statistisches Wissen aufeinandertreffen und in Zusammenhang gebracht werden müssen, um die menschliche Fähigkeit nachzubilden, Bedeutungen aus dem Kontext heraus zu verstehen.

Ansätze bieten hier unter anderem Bays'sche Netze (dienen zur Repräsentation von und zum Schließen bei unsicherem Wissen), das Markov Modell (ermöglichen es, **aus Indizien auf die tatsächlichen Vorgänge im Hintergrund zu schließen**) oder das Case Base Reasoning (fallbasiertes Schließen erlaubt das Lösen von Problemen mithilfe von Erfahrungswissen, das in Form von Fällen in einer Fallbasis gespeichert ist).

---

6 <https://www.nature.com/news/can-we-open-the-black-box-of-ai-1.20731>, Can we open the black box of AI? Artificial intelligence is everywhere. But before scientists trust it, they first need to understand how machines learn by Davide Castelvecchi, 05 October 2016 published in Nature – the International Journal of Science.

7 <https://heise.de/-3695745>, Bilderkennungsprogramme von autonomen Fahrzeugen können ausgetrickst werden, Florian Rotzer, 07. Mai 2017.

Einen Überblick bietet das Schaubild

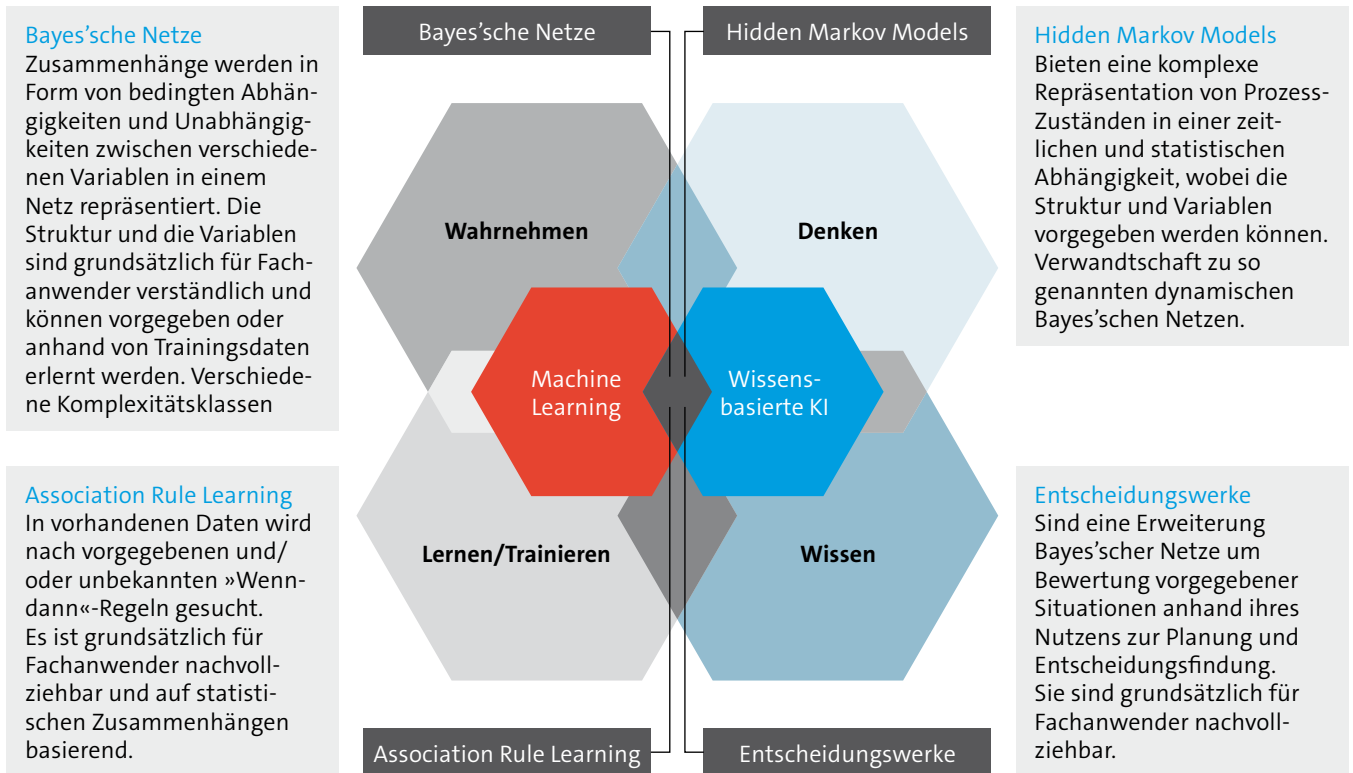


Abbildung 6: Hybride KI: Verknüpfung zwischen statistischen Lernmethoden und großen Wissensrepräsentationen (subsymbolischer KI), adaptiert von intelligent views GmbH

In einer hybriden KI werden Wissen und Daten wertschöpfend zusammengeführt und eine »explainable AI« geschaffen.



## 3 Use Case: HR/Recruiting

### Use Case

Automatisierte Entscheidungsverfahren (ADM-Verfahren) werden immer häufiger im Recruiting und dem Personalmanagement eingesetzt, u. a. für die folgenden Aufgaben:

- Sourcing von geeigneten Kandidaten durch Datenanalyse in sozialen Medien oder auf Internet-Plattformen
- Erstbewertung und Vorsortierung von Lebensläufen/Bewerbungsdokumenten z. B. mithilfe automatisierter Textverarbeitung oder anhand strukturierter Daten
- Entscheidungsunterstützung bei der Bewertung im Rahmen des Bewerbungsprozesses, z. B. mithilfe von maschinellen Lernverfahren welche historische Bewerbungsdaten als Trainingsgrundlage erhalten.
- Kontinuierliche Bewertung und Evaluation von Mitarbeitern oder Teams mithilfe datengestützter Verfahren und maschinellem Lernen (»people analytics«)

Beim Einsatz solcher Systeme sind in Deutschland eine Vielzahl von rechtlichen Rahmenbedingungen zu beachten, u. a. die DS-GVO, das Bundesdatenschutzgesetz (BDSG-Neu), das Allgemeine Gleichbehandlungsgesetz (AGG) sowie das Betriebsverfassungsgesetz (BetrVG).

Ein Risiko beim Einsatz komplexer ADM-Verfahren ist die nicht immer gegebene Nachvollziehbarkeit und Transparenz von Entscheidungen die mithilfe dieser Systeme getroffen werden. Zusätzlich besteht das Risiko, dass ADM-Verfahren Probleme aus bestehenden Prozessen wie z. B. Diskriminierung quasi automatisieren, indem sie diese anhand von Trainingsdaten »erlernen« und reproduzieren. Und nicht zuletzt unterliegt die Verarbeitung personenbezogener Daten natürlich selbst strengen Datenschutzvorschriften, was gerade beim Einsatz von Machine Learning Probleme verursachen kann, da beispielsweise viele Modelle und Verfahren die Wahrung der Anonymität von Trainingsdaten nicht garantieren können.

Bereits heute ist daher eine Vielzahl von Anforderungen beim Einsatz solcher Verfahren zu beachten. Im Rahmen einer Case-Study zeigen wir in den folgenden Abschnitten exemplarisch, welche Herausforderungen beim Einsatz von ADM-Verfahren im HR-Bereich entstehen und wie wir diesen begegnen können.

### 3.1 Problemdefinition

Ein Unternehmen möchte mithilfe von ADM-Verfahren und Machine Learning die Vorauswahl und Bewertung von Kandidaten im Bewerbungsprozess verbessern. Ziel ist, das Verfahren zu beschleunigen und die Kündigungsquote erfolgreicher Kandidaten zu verringern.

## 3.2 Risiko- und Nutzenbewertung sowie Zieldefinition

Zunächst sollten Risiken und Nutzen des Verfahrens bewertet werden. Im Idealfall liefert die Nutzenbewertung hierbei eine grobe quantitative Abschätzung der zu erwartenden Effizienzgewinne sowie sonstiger auch qualitativer Vorteile die bei der Einführung des ADM-Verfahrens zu erwarten sind. Diese könnten in dem hier behandelten Fall z. B. sein:

- Reduktion der Dauer eines Bewerbungsverfahrens um 30 % durch schnellere Rückmeldungen und Automatisierung einzelner Entscheidungen
- Steigerung der Zufriedenheit von Mitarbeitern durch eine bessere Unterstützung beim Recruiting neuer Kollegen
- Steigerung von Objektivität, Nachvollziehbarkeit und Reproduzierbarkeit bei der Auswahl von Bewerbern
- Bessere Möglichkeiten zur quantitative Analyse von Entwicklungszielen wie z. B. der Förderung von Diversität
- Die Fähigkeit, strukturelle Probleme wie Diskriminierung in Bewerbungsprozessen zu entdecken und zu beseitigen und so Compliance-Risiken zu senken

Anhand der Nutzenanalyse sollten auch quantitativ oder qualitativ verifizierbare Zielkriterien definiert werden, welche im nächsten Schritt bei der Umsetzung des Verfahrens zum Training und zur Kontrolle des ADM-Verfahrens genutzt werden können. In diesem Beispiel könnte dies z. B. die durchschnittliche Dauer eines Bewerbungsprozesses, die Bewertung des Verfahrens durch die Bewerber oder die durchschnittliche Beschäftigungsdauer erfolgreicher Bewerber im Unternehmen sein. Ein einzelnes messbares Bewertungskriterium zu finden das die definierten Ziele perfekt abbildet ist dabei fast nie möglich. Es sollten daher besser mehrere, sich ergänzende Kriterien definiert werden, wobei auch qualitative Metriken wie schriftliches Feedback herangezogen werden können.

Dem Nutzen gegenüber stehen die Risiken beim Einsatz von ADM-Systemen, welche bereits im Vorfeld eines Projektes analysiert werden sollten und für den hier behandelten Fall u. a. wären:

- Mangelnde Nachvollziehbarkeit von algorithmischen Entscheidungen die bei der Auswahl oder der Bewertung von Bewerbern eingesetzt werden
- Mögliche Automatisierung von bestehenden Problemen durch mangelhafte oder ungeeignete Trainingsdaten
- Verschlechterung der Qualität der (Vor-)Auswahl von Bewerbern durch den Einsatz ungeeigneter Algorithmen
- Unzufriedenheit bei Bewerbern aufgrund mangelnder Nachvollziehbarkeit oder dem unpersönlichen Charakter von automatisierten Entscheidungen
- Höhere Kosten durch den Einsatz von wartungsintensiven technologischen Lösungen
- Verlust von Steuerungsfähigkeit aufgrund mangelnder Mechanismen zur gezielten Anpassung des ADM-Systems
- Steigerung des Risikos von Datenverlust oder Datenmissbrauch durch die automatisierte Verarbeitung von Bewerberdaten

Oft ist die Anzahl der vorstellbaren Risiken wie hier bei der Analyse länger als die Liste der Vorteile, was jedoch nicht heißen muss, dass der Einsatz eines ADM-Verfahrens nicht sinnvoll ist und den Risiken nicht durch geeignete Maßnahmen begegnet werden kann. In der Praxis zeigen sich zudem bei der Umsetzung der weiteren Schritte oft noch weitere Vorteile und Risiken, die bei der Analyse übersehen oder nicht dokumentiert wurden, jedoch ebenfalls bewertet werden sollten.

### 3.3 Auswahl geeigneter Daten zum Training des Systems

Nachdem Ziele und Risiken bei der Einführung des ADM-Verfahrens analysiert wurden müssen geeignete Daten identifiziert werden, die als Grundlage für das Verfahren dienen können. Hier sollten interne wie externe Datenquellen in Betracht gezogen werden. Für jeden Datentyp sollte dann anhand **rechtlicher**, **organisatorischer** und **ethischer** Grundsätze analysiert werden, ob eine Erhebung und Nutzung möglich und sinnvoll ist und welche Voraussetzungen gegebenenfalls bestehen. Für den hier behandelten Fall kommen u. a. folgende Datenquellen in Betracht:

- Von Bewerbern bereitgestellte Informationen: Hier ist zwischen strukturierten Daten wie z. B. über Formulare bereitgestellten Informationen über Abschlüsse und Berufserfahrung sowie unstrukturierten Daten wie z. B. Anschreiben und frei formatierten Lebensläufen zu unterscheiden
- Zusätzliche Informationen über Bewerber die aus öffentlichen Quellen stammen, wie z. B. Profil-Daten von Online-Plattformen
- Historische Daten und Statistiken zu Bewerbern und Angestellten aus unternehmens-internen IT-Systemen
- Kontextinformationen die zur Anreicherung von Daten genutzt werden können, z. B. Ranking-Informationen über Hochschulen oder angegebene Drittunternehmen
- Personenbezogene Daten zu geschützten Merkmalen wie Herkunft, Geschlecht und Alter, welche ggf. benötigt werden um Transparenz- und Fairnessanforderungen sicherzustellen und zu überwachen

Insbesondere personenbezogene Daten können nur genutzt werden, wenn geeignete Sicherheits- und Vorsorgemaßnahmen getroffen werden, wie z. B. die Pseudonymisierung der Daten. Insbesondere die Erhebung sensibler personenbezogener Daten wie Geschlecht und Alter von Bewerbern mag hier absurd erscheinen, diese Daten sind aber notwendig, um beispielsweise Diskriminierung von Bewerbern gezielt verhindern zu können.

Zusätzlich sollte analysiert werden, ob und wie die definierten Zielkriterien mit den vorhandenen Daten überhaupt abgebildet werden können, ob hierfür zusätzliche Daten erhoben werden müssen oder ob eine Umsetzung anhand der vorhandenen Daten überhaupt möglich ist. Insbesondere der letzte Fall wird in der Praxis leider oft ignoriert, was dazu führt, dass ADM-Vorhaben umgesetzt werden ohne eine realistische Chance auf die Realisierung oder Messung der zu erreichenden Ziele zu haben.

### 3.4 Untersuchung der Datenqualität und möglicher Probleme

Nachdem man geeignete Datenquellen identifiziert und ausgewählt hat, kann eine exemplarische Erhebung der Daten erfolgen, welche auch manuell und mit gegenüber dem fertigen Verfahren höherem Aufwand erfolgen kann. Die so gewonnenen Beispieldaten können dann auf ihre Qualität und Eignung untersucht werden, indem z. B. statistische Auswertungen erfolgen.

Geeignete Fragen sind hierzu u. a.:

- Gibt es einen messbaren Zusammenhang der betrachteten Daten mit den definierten Zielgrößen (in diesem Beispiel z. B. die durchschnittliche Beschäftigungsdauer erfolgreicher Bewerber)? Dies kann z. B. über eine Korrelationsanalyse oder die einfache Untersuchung relativer Häufigkeiten analysiert werden. Daten, die keinerlei statistisch nachweisbaren Zusammenhang zu Zielgrößen aufweisen, können dann von der Betrachtung ausgeschlossen werden.
- Weisen die Daten einen messbaren Zusammenhang mit bestimmten geschützten Merkmalen wie Alter oder Geschlecht auf? Falls ja sollte ein geeignetes Verfahren gewählt werden um sicherzustellen, dass durch den Einsatz der Daten im ADM-Prozess keine unzulässige Diskriminierung von Bewerbern erfolgen kann.
- Wie ist Qualität und Zuverlässigkeit der erhobenen Daten? Bestehen zu viele Unregelmäßigkeiten oder Fehler in den Daten kann die Verwendung im ADM-Prozess problematisch sein, da fehlerhafte Daten die Korrektheit und Zuverlässigkeit des eingesetzten Verfahrens reduzieren können.
- Können die Daten überhaupt in geeigneter Weise für ein ADM-Verfahren aufbereitet werden? Insbesondere bei unstrukturierten Daten wie Freitexten, aber auch bei kategorialen oder numerischen Daten müssen diese zunächst für das eingesetzte Verfahren passend aufbereitet werden. Falls dies nicht sinnvoll möglich ist können sie entsprechend nicht genutzt werden.
- Können die Daten langfristig gespeichert und für das Training des ADM-Verfahrens verwendet werden? Insbesondere bei personenbezogenen Daten besteht fast immer eine Zweckbindung die eine unbefristete Speicherung und Verarbeitung untersagt. Diese Daten müssen daher gegebenenfalls anonymisiert werden um sie weiter nutzen zu können, oder nach einer gegebenen Zeit aus dem System entfernt werden.

Durch den Analyseprozesses sollte eine erste quantitative Bewertung der Datenqualität und -eignung für das geplante ADM-Verfahren geschaffen sein. Dies hilft im nächsten Schritt, ein geeignetes Verfahren auszuwählen und umzusetzen. In der Praxis sind zwischen Datenerhebung und -analyse einerseits und der Auswahl und Umsetzung eines Verfahrens andererseits jedoch oft mehrere Iterationen nötig um ein funktionierendes System zu erhalten.

### 3.5 Umsetzung eines Machine Learning Verfahrens

Nachdem eine erste Datenauswahl und -erhebung erfolgt ist kann ein geeignetes Verfahren ausgewählt werden, um die Daten im ADM-Prozess zu verarbeiten und Vorhersagen zu treffen. Mögliche Verfahren reichen von einfachen statistischen oder heuristischen Methoden bis hin zu sehr komplexen Verfahren wie dem aktuell sehr gefragten »Deep Learning« Ansatz. Dieses wird dann mit den erhobenen Daten darauf trainiert, die gewählten Zielgrößen anhand der Eingabedaten vorherzusagen. Zur Beurteilung der Qualität und Eignung der Ergebnisse werden dann typischerweise statistische Tests und Metriken herangezogen, die beispielsweise die Anzahl falsch vorhergesagter Testdaten auswerten. Die erreichte Genauigkeit des Verfahrens entscheidet dann oft über dessen Eignung für den Anwendungsfall. Im vorliegenden Fall würde das Verfahren beispielsweise versuchen, anhand der Daten über einen gegebenen Bewerber vorzusagen ob dieser eingestellt wurde oder nicht.

### 3.6 Evaluierung und kontinuierliche Überwachung der Ergebnisse

Um die Qualität des Verfahrens bewerten und überwachen zu können, sollten die vorher definierten Zielkriterien kontinuierlich erfasst und ausgewertet werden. Im einfachsten Falle kann die Genauigkeit (accuracy) gemessen werden, indem die Anzahl der »korrekt« und »inkorrekt« klassifizierten Datenpunkte erfasst wird. Gerade bei HR-Entscheidungen sind hierfür spezielle Verfahren erforderlich: Soll z. B. die Korrektheit von Einstellungsentscheidungen gemessen werden indem geprüft wird welche Kandidaten nach 6 Monaten noch im Unternehmen arbeiten, so können ohne Anpassungen des Verfahrens lediglich Fehler 1. Art (ein Kandidat wird eingestellt aber entspricht nicht den Anforderungen), nicht aber Fehler 2. Art (ein Kandidat wird nicht eingestellt aber entspräche den Anforderungen) erfasst werden. Insbesondere um die Korrektheit und nötige Sorgfalt beim Einsatz des Verfahrens nachzuweisen sollten jedoch beide Fehlerarten untersucht werden. Techniken wie z. B. randomisierte Tests können hierbei hilfreich sein. Generell sollten Modellvorhersagen und Eingabedaten in regelmäßigen Abständen geprüft und untersucht werden, um die Eignung und Qualität des Modells und der Daten zu überwachen.

### 3.7 Datenschutzrechtliche Anforderungen

Im angeführten Beispiel werden verschiedene personenbezogene Daten im Rahmen des ADM-Verfahrens eingesetzt. Um diese Daten nutzen zu können, müssen verschiedene Aspekte der DS-GVO betrachtet werden, insbesondere hinsichtlich der Transparenz, Informationspflichten und Rechtsgrundlagen. Insbesondere die grundsätzlichen Transparenzanforderungen sind im Hinblick auf Machine Learning interessant, da – wie oben gezeigt – nicht alle Vorgänge immer abschließend vorab erklärbar sind. Da die Transparenzregeln auch die Information über die für die Datenverarbeitung notwendige Rechtsgrundlage voraussetzen, soll im Kapitel 4 zudem erläutert werden, welche Rechtsgrundlagen für Datenverarbeitung in Machine Learning Anwendungen in Betracht kommen. Mit den einzelnen zu erfüllenden Informationspflichten der DS-GVO beschäftigt sich ein separater Bitkom Leitfaden, in dem dann auch die Anforderungen aus dem Informationspflichten-Katalog der DS-GVO näher beleuchtet werden sollen.<sup>8</sup>

### 3.8 Weitere rechtliche Anforderungen und Methoden zur Umsetzung

Neben der DS-GVO ist für den gezeigten Anwendungsfall eines automatisierten Entscheidungsverfahrens auch das Allgemeine Gleichbehandlungsgesetz (AGG) relevant, welches eine Benachteiligung aufgrund verschiedener Merkmale wie des Geschlechts, der Religion oder Weltanschauung verbietet. Zusätzlich können Mitbestimmungsgesetze wie das Betriebsverfassungsgesetz (BetrVG) oder das Bundesbeamtengesetz (BBG) relevant sein.

Um eine Diskriminierung anhand der im AGG genannten geschützten Attribute (Geschlecht, Religion, Alter, ...) im Rahmen eines automatisierten Entscheidungsverfahrens zu vermeiden, müssen zunächst die entsprechenden Attributwerte der betroffenen Personen erfasst werden. Die Speicherung dieser sensitiven Daten sollte dabei strikt getrennt und nach höchsten Sicherheitsstandards erfolgen. Mit diesen Daten kann anschließend untersucht werden, ob durch das eingesetzte Verfahren möglicherweise eine Diskriminierung anhand dieser Attribute erfolgt. Hierzu existieren verschiedene statistische Ansätze, »Disparate Impact« hat sich als statisches Maß zumindest im amerikanischen Raum als Standard hierfür etabliert. Zur Bestimmung dieses Maßes wird zunächst die Erfolgsquote von Personen gemessen, die den untersuchten Auswahlprozess durchlaufen haben, wobei Personen anhand der Werte eines sensitiven Attributes – z. B. Geschlecht – aufgeteilt werden. Anschließend wird das Verhältnis der Erfolgsquoten für die einzelnen Gruppen berechnet. Fällt dieses unter einen gegebenen Wert (z. B. 80 %) kann darauf geschlossen werden, dass eventuell eine Diskriminierung vorliegt (unter Vorbehalt anderer statistischer Zusammenhänge).

---

<sup>8</sup> Veröffentlichung voraussichtlich Ende 2018.

Ein Beispiel: Durchlaufen z. B. 20 % der Männer aber nur 10 % der Frauen einen Auswahlprozess erfolgreich, ergibt sich ein Wert von 0.5 (oder 50 %) für den Disparate Impact, was auf eine starke Diskriminierung hindeuten kann. Die so identifizierten Fälle können dann anschließend im Detail untersucht werden um festzustellen wodurch sich das statistische Ungleichgewicht möglicherweise ergibt. Disparate Impact ist ein sehr einfaches Fairnessmaß und kann nicht sämtliche Arten von Diskriminierung zuverlässig identifizieren, dementsprechend sollte es nicht als einzige Kennzahl bei der Bewertung von Entscheidungsverfahren eingesetzt werden.

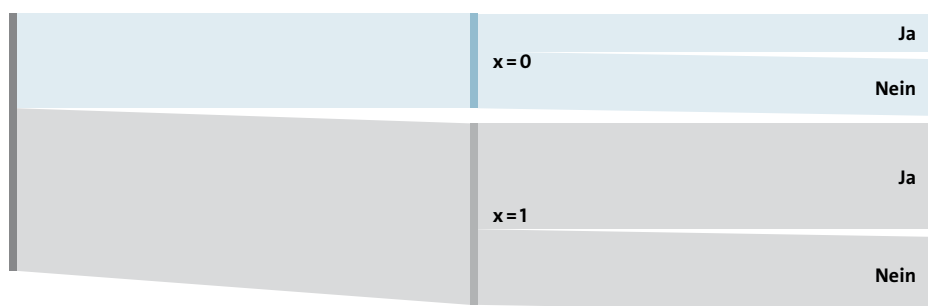


Abbildung 7: Entscheidungsbaum für einen Prozess der von zwei Gruppen von Personen – z. B. Männern und Frauen – durchlaufen wird, wobei für jede Gruppe die Wahrscheinlichkeit gemessen wird, den Prozess erfolgreich oder nicht erfolgreich zu durchlaufen. Aus dem Verhältnis der gruppenspezifischen Erfolgswahrscheinlichkeit kann die Kenngröße des »Disparate Impact« berechnet werden. Adaptiert von 7scientists

Neuere Methoden zur Messung und Gewährleistung von Fairness bei automatisierten Entscheidungsprozessen wie z. B. »Fairness Through Awareness«<sup>9</sup> versuchen, Fairness anhand der Ähnlichkeit von Datenpunkten zu bewerten: Die Annahme hierbei ist, dass ein Verfahren zu ähnlichen Entscheidungen bezüglich zweier Personen gelangen sollte, wenn sich die relevanten Daten der beiden Personen stark ähneln. Hierbei muss zur Messung der Ähnlichkeit ein geeignetes mathematisches Maß verwendet werden, welches nur legitime Attribute heranzieht, also z. B. Merkmale wie Geschlecht oder Alter nicht einbezieht. Eine Weiterentwicklung des Verfahrens<sup>10</sup> bietet zusätzlich einen Weg, um aus einem Ursprungsdatensatz, der unfaire oder nicht zulässige Attribute enthält, einen Datensatz mit einer fairen Repräsentation zu generieren, der anschließend mit einem beliebigen Verfahren weiterverarbeitet werden kann. Beide Verfahren haben eine Reihe an Voraussetzungen die nicht immer einfach zu erfüllen oder umzusetzen sind, dementsprechend sind sie bisher nicht uneingeschränkt in der Praxis nutzbar.

9 <https://arxiv.org/abs/1104.3913>.

10 <https://www.cs.toronto.edu/~toni/Papers/icml-final.pdf>.

# 4 Machine Learning und die Datenschutzgrundverordnung

Die vorangegangenen Kapitel haben einen Überblick über die technischen Spezifika gegeben und die praktische Anwendung des Machine Learning erläutert. Deutlich ist dabei vor allem eines geworden: Machine Learning ohne Daten ist ungefähr gleichzusetzen mit einer Mühle ohne Wasser. Datenverarbeitungen, und das ist nicht erst seit der viel besprochenen DS-GVO der Fall, unterliegt strengen Regeln, wenn die Daten einen Personenbezug aufweisen. In diesem Kapitel soll daher aufgezeigt werden, welche Anforderungen aus dem Datenschutzrecht sich für Machine Learning ergeben und wie diese erfüllt werden können. Die DS-GVO sollte dabei keineswegs nur als Restriktion oder gar Hemmnis verstanden werden, bietet sie doch, vor allem durch ihre strengen Anforderungen, einen der wichtigsten Rahmen für das gesellschaftliche Bedürfnis nach Schutz der Person hinter den Daten.

Grundsätzlich gilt zunächst folgendes: Datenverarbeitungen unter Einsatz von ML oder KI fallen, soweit personenbezogene Daten betroffen sind, in den Anwendungsbereich der DS-GVO.<sup>11</sup> Das erfasst natürlich auch die Einhaltung der Transparenzgrundsätze (insbesondere Informationspflichten), sowie die Geltung des Verbots mit Erlaubnisvorbehalt. Aufgrund der erläuterten Funktionsweisen und Lernvorgänge verdient dies eine genauere Untersuchung, da sich insbesondere die Frage stellt, wie eine adäquate Information des Nutzers trotz der möglichen Unkenntnis des Anwenders hinsichtlich des konkret ablaufenden Lernprozesses ausgestaltet werden kann.

## 4.1 Transparenzgrundsatz und Informationspflichten

Transparenz der Verarbeitung ist einer der Grundsätze der DS-GVO (Art. 5 Abs. 1 lit a DS-GVO) und regelt, dass personenbezogene Daten in einer für den Betroffenen nachvollziehbaren Art und Weise verarbeitet werden müssen. Dies umfasst sowohl die Informationen über den Verarbeitungsvorgang zum Zeitpunkt der Erhebung, als auch z. B. die Rechenschaftspflicht (Art. 5 Abs. 2 DS-GVO), die bestimmt dass der Verantwortliche die Einhaltung der Regelungen des Art. 5 Abs. 1 DS-GVO nachweisen können muss. Die Bestimmung ist im Zusammenhang mit Art. 24 Abs. 1 DS-GVO zu lesen, welcher dem Verantwortlichen die Umsetzung geeigneter technischer und organisatorischer Maßnahmen aufgibt, welche u. a. auch Nachweiszwecken zu dienen haben.

Der Kernbereich der Transparenzanforderungen wird von Art. 12 ff. DS-GVO ausgefüllt. Art. 12 DS-GVO regelt zunächst vor die Klammer gezogen, dass alle Informationen und Mitteilungen in präziser, transparenter, verständlicher und leicht zugänglicher Form in einer klaren und einfachen Sprache zu erfolgen haben. Die Informationen sind gem. Art. 12 Abs. 5 DS-GVO grundsätzlich unentgeltlich zur Verfügung zu stellen.

---

<sup>11</sup> Eine umfassende Erläuterung zum Begriff des »personenbezogenen Datums« findet sich im Bitkom Faktenpapier zu Blockchain und Datenschutz im Kapitel 3.1.1.: <https://www.bitkom.org/noindex/Publikationen/2018/Leitfaeden/Blockchain-und-Datenschutz/180502-Faktenpapier-Blockchain-und-Datenschutz.pdf>.



## 4.2 Wie muss über Zwecke der Verarbeitung belehrt werden?

Die Zweckbindung einer jeden Verarbeitung personenbezogener Daten ist ein tragender Grundsatz der DS-GVO.<sup>12</sup> Sie unterteilt sich in das Gebot der

- Zweckfestlegung, wonach eine Verarbeitung nur für festgelegte, eindeutige und legitime Zwecke erfolgen darf und die
- Zweckbindung im engeren Sinne, dem Verbot der Verarbeitung personenbezogener Daten in einer Weise, die mit dem Erhebungszweck unvereinbar ist.<sup>13</sup>

Soweit Daten beim Betroffenen erhoben werden (Direkterhebung), sieht Art. 13 Abs. 1 lit. c DS-GVO vor, dass dem Betroffenen die verfolgten Zwecke der Verarbeitung der konkreten Daten mitzuteilen sind.

Eine Ausnahme zu diesem Grundsatz findet sich Art. 13 Abs. 4 DS-GVO, nach dem keine Mitteilung erfolgen muss, soweit der Betroffene bereits über die Informationen verfügt.

Zu klären ist zunächst, in welcher Form die Mitteilung zu erfolgen hat. Formvorgaben folgen aus Art. 12 Abs. 1 S. 1 DS-GVO. Hiernach ist in präziser, verständlicher und leicht zugänglicher<sup>14</sup> Form, sowie in klarer und einfacher Sprache zu informieren.

Von großer praktischer Bedeutung ist die Frage, wie detailliert über die verfolgten Zwecke zu informieren ist. Art. 5 Abs. 1 lit. b DS-GVO statuiert insoweit, dass die Zwecke festgelegt, eindeutig und legitim sein müssen. Die Tatsache, dass eine sehr enge Festlegung der Zwecke schneller zur Erforderlichkeit einer Zweckänderung (Art. 6 Abs. 4 DS-GVO) und damit zu mehr Aufwand oder der Unmöglichkeit einer rechtmäßigen Weiternutzung der Daten führen kann, verleiht der Frage zusätzliche Relevanz.<sup>15</sup>

Die DS-GVO liefert auf die Frage nach den Präzisionsanforderungen bzw. darauf, wie abstrakt ein Zweck noch beschrieben werden darf, keine Antwort.<sup>16</sup>

---

12 Dammann, Erfolge und Defizite der EU-Datenschutzgrundverordnung – Erwarteter Fortschritt, Schwächen und überraschende Innovationen, in: ZD 2016, 307 (311).

13 BeckOK DatenschutzR / Schantz, DS-GVO, Art. 5 Rn. 12; Schantz/Wolff, Das neue Datenschutzrecht, Rn. 400.

14 S. hierzu Walter, Die datenschutzrechtlichen Transparenzpflichten nach der Europäischen Datenschutz-Grundverordnung, in: DSRI TB 2016, 367 (368 f.), welcher vorschlägt, »leicht zugänglich« im Sinne von »gedanklich leicht zugänglich«, also als »einfach« zu verstehen.

15 Veil, DS-GVO: Risikobasierter Ansatz statt rigides Verbotprinzip – Eine erste Bestandsaufnahme, in: ZD 2015, 347 (349).

16 Paal/Pauly/Hennemann, DSGVO, Art. 13 Rn. 5a.

### 4.3 Artikel 6 Absatz 4 und noch unbekannte Nutzungszwecke

Eine Belehrungspflicht über die Verarbeitungszwecke folgt aus Art. 13 Abs. 1 lit. c Alt. 1, 14 Abs. 1 lit. c Alt. 1 DS-GVO.

Eine konkrete und präzise Belehrung über noch unbekannte Nutzungszwecke ist logischerweise nicht möglich, dennoch ist eine Weiterverarbeitung von Daten zu mit dem ursprünglichen Zweck kompatiblen Zwecken unter den Voraussetzungen des Art. 6 Abs. 4 DS-GVO möglich.

Es muss daher einen Weg geben zwischen einer hinreichend konkreten Belehrung, die jedoch u. U. zu einem späteren Verlust der Nutzbarkeit der Daten für andere Zwecke führen kann und einer zu abstrakten – und damit ggf. nicht mehr rechtskonformen – Belehrung, welche noch unbekannte Zwecke offen hält, oder die Nachinformationen entsprechend zu geben, wenn Art. 6 Abs. 4 DS-GVO greift.

In jedem Fall ist zu versuchen, die – auch künftigen – Zwecke so konkret wie möglich zu umschreiben. Künftige denkbare Zwecke können etwa unter Bemühung der Kreativität des Verfassers in Form von Regelbeispielen aufgeführt werden.

Dem Betroffenen ist gem. Art. 13 Abs. 1 lit. c Alt. 2, 14 Abs. 1 lit. c Alt. 2 DS-GVO bei Erhebung auch die Rechtsgrundlage der Datenverarbeitungen mitzuteilen.

Die möglichen Rechtsgrundlagen finden sich in Art. 6 Abs. 1 DS-GVO. Flankiert werden sie vom vor die Klammer gezogenen und stets mitzulesenden Art. 5 Abs. 1 lit. b DS-GVO, welcher hinsichtlich der Zweckbindung vorschreibt, dass eine Weiterverarbeitung erhobener Daten nur erfolgen darf, sofern sie mit den ursprünglichen Zwecken vereinbar ist.

Festzustellen ist also, dass eine Weiterverarbeitung bereits vorhandener personenbezogener Daten zu geänderten Zwecken nur erfolgen darf, wenn

- eine Rechtsgrundlage i. S. d. Art. 6 Abs. 1 gegeben ist und
- der neue Zweck mit dem ursprünglich Zweck vereinbar ist.

Das Vorliegen einer Rechtsgrundlage unterstellt, ist zu untersuchen, ob Art. 6 Abs. 4 DS-GVO<sup>17</sup> für Machine Learning verwertbare Ansätze enthält.

Art. 6 Abs. 4 DS-GVO enthält einen nicht abschließenden Katalog mit Kriterien zur Beurteilung der Kompatibilität von Erhebungszweck und Weiterverarbeitungszweck, wobei für die Beurteilung die Perspektive des Verantwortlichen maßgeblich ist.<sup>18</sup>

<sup>17</sup> Dessen Rechtsnatur bereits nicht eindeutig ist, vgl. BeckOK DatenschutzR/Albers/Veit, DS-GVO, Art. 6 Rn. 71 ff.

<sup>18</sup> BeckOK DatenschutzR/Albers/Veit, DS-GVO, Art. 6 Rn. 69.

Zwischenergebnis: Es bietet sich an, sofern die ursprüngliche Datenverarbeitung auf einem Vertragsverhältnis nach Art. 6 Abs. 1 lit. b DS-GVO beruht, die Zwecke in diesem Verhältnis transparent aber hinreichend flexibel zu formulieren, um zu späteren Zeitpunkten eine Vereinbarkeit mit den neuen Zwecken erzielen zu können.

## 4.5 Rechtsgrundlagen

### 4.5.1 Datenverarbeitung für ML auf Basis der Rechtsgrundlage der Vertragserfüllung

Die Verarbeitung personenbezogener Daten für ML-Zwecke kann auf verschiedene in der DS-GVO genannte Rechtsgrundlagen gestützt werden. Namentlich kommen die Vertragserfüllung, das berechnete Interesse und die Einwilligung als Rechtsgrundlage in Betracht. Entgegen einem weit verbreiteten Missverständnis sind Einwilligung und berechtigtes Interesse weder die einzigen noch die häufigsten herangezogenen Rechtsgrundlagen. Eine breite Palette von Verarbeitungen personenbezogener Daten für ML kann auf den Erlaubnistatbestand der Vertragserfüllung gestützt werden.

### 4.5.2 Rechtliche Grundlagen

Art. 6 Abs. 1 lit. b DS-GVO rechtfertigt eine Verarbeitung, soweit sie »für die Erfüllung eines Vertrages erforderlich ist, dessen Vertragspartei die betroffene Partei ist [...]«, d. h. es muss ein direkter Zusammenhang zwischen der Verarbeitung der Daten einer Person und der Erforderlichkeit, eine bestimmte vertragliche Leistung an diese Person zu erbringen, bestehen.

In vielen Fällen ist die Verarbeitung personenbezogener Daten für Machine Learning Voraussetzung für die Erreichung der Vertragserfüllung, d. h. im Rahmen des ursprünglichen Vertragszwecks, auch wenn in diesem Rahmen Daten im Rahmen des Lernprozesses für neue Zwecke weiterverarbeitet werden. In diesen Fällen wird ML verwendet, um die Leistungserbringung an die Person zu ermöglichen, deren Daten verarbeitet werden. Der Leistungsanspruch der Verarbeitung personenbezogener Daten für ML wird nicht ausschließlich durch die Beurteilung bestimmt, ob eine Leistung überhaupt erbracht werden kann oder nicht – vielmehr ist es entscheidend, ob sie in der vertraglich vereinbarten Form und Qualität erbracht werden kann. Zu diesem Zweck enthalten ML-basierte Vertragsleistungen oft ein dynamisches oder proaktives Element, für das der Nutzer Verträge abschließt und das ohne die Erkenntnisse aus ML nicht durchgeführt werden könnte.

### 4.5.3 Leistungen, bei denen ML in vertraglich vereinbarter Art und Weise funktionieren muss

ML ist nicht völlig neu und für historischere ML-basierte Produkte und Dienstleistungen ist ihr Einsatz und ihre Notwendigkeit im Rahmen der vertraglichen Leistungsanforderungen nie in Frage gestellt worden. Beispielsweise verfügen viele Autos über Automatikgetriebe, die die persönlichen Vorlieben des Fahrers kennenlernen und die Schaltung entsprechend anpassen.

Im Falle einer fehlerhaften Schaltung oder um die Leistung im Allgemeinen zu verbessern, werden neuere Fahrzeuge vernetzt, um dem Hersteller Lerndaten zur Verfügung zu stellen, und bei älteren Fahrzeugen werden die Erkenntnisse häufig ausgelesen, um Leistungsprobleme zu lösen. Die damit verbundene Verarbeitung personenbezogener Daten ist zur Erfüllung einer vertraglichen Lieferverpflichtung und im Falle des Leasings zur Aufrechterhaltung eines mangel-freien Produktes erforderlich. Es gibt in diesem Fall keine Umnutzung der personenbezogenen Daten für ML; das Lernen ermöglicht vielmehr die Leistungserbringung, da ohne den Lernvorgang kein adaptives Automatikgetriebe angeboten werden könnte.

Gleiches gilt für alle anderen Arten der Vertragserfüllung, die ML zur ordnungsgemäßen Erfüllung benötigen. Der Bedarf wird durch die vereinbarte Leistung bestimmt. Ist für die vereinbarte Leistung in der zugesicherten Qualität die Verarbeitung personenbezogener Daten über ML erforderlich, so ist dies auch nach Art. 6 Abs. 1 lit. b DS-GVO gerechtfertigt.

Dies gilt z. B. für Dienste wie die natürliche Spracherkennung, die äußerst komplex sind und unabhängig vom Entwicklungsaufwand im Vorfeld und dem Testen von Beta-Nutzergruppen stets ein mangelndes Verständnis z. B. für regionale Besonderheiten wie Dialekte oder Akzente aufzeigen, die der Nutzer als Teil seiner natürlichen Sprache erwartet, die das System aber erst mit der Zeit lernen kann. Im Rahmen der entsprechenden Verträge erwarten die Nutzer auch, dass die natürlichen Spracherkennungssysteme mit der Sprachentwicklung Schritt halten. Dies ist durchaus eine Herausforderung: Neue Wörter oder Phrasen werden populär oder werden erfunden – manchmal über Nacht. Mikro-Dialekte entstehen und passen sich an. Die meisten Sprachverständnis-Engines sind auf Machine Learning angewiesen, um sich an sich ändernde Sprachmuster in Echtzeit anzupassen. Kunden, die Dienste zum Verstehen der gesprochenen Sprache (spoken language understanding, »SLU«) nutzen, erwarten, dass diese Dienste auch dann funktionieren, wenn Sprachmuster auftreten oder sich ändern. Sie bestimmen die »natürliche Sprache« und wären frustriert, wenn sie warten müssten, bis ein neues Muster von Sprachwissenschaftlern manchmal Jahre später offiziell anerkannt wird. Die Erbringung von SLU-Dienstleistungen erfordert daher ein proaktives Element, das nur ML liefern kann. Genau das macht ML daher zu einer vertraglichen Anforderung.

#### **4.5.4 Fehler- und Mängelbeseitigung, um die Leistung auf dem vertraglich vereinbarten Niveau zu halten**

Aus rechtlicher Sicht könnte die fehlerhafte, unkorrekte Spracherkennung sogar als rechtlich relevanter Mangel eingestuft werden, was den Anbieter der entsprechenden Geräte oder Dienstleistungen wiederum verpflichten würde, die Mängel zu beheben. Dazu ist es erforderlich, die gesprochenen Eingaben des Kunden zu verarbeiten und zu überprüfen, ob diese verstanden werden und aus Missverständnissen zu lernen, um deren Wiederholung zu vermeiden. Solche Verbesserungen können nicht für jeden Kunden einzeln vorgenommen werden, wie es nach Art. 6 Abs. 1 lit. b DS-GVO oft formell gefordert wird, da die Leistungspflicht gegenüber der Vertragspartei bestehen muss, deren Daten verarbeitet werden, um eine solche Leistung zu ermöglichen. Dies auf Kundenbasis zu tun, nur um die direkte Verbindung zwischen den verarbeiteten Daten und den erbrachten Dienstleistungen für dieselbe Person aufrechtzuerhalten, würde zu einer redundanten Fehlererkennung und -behebung führen. Dies wäre ineffektiv und langsam und würde mehrere Benutzer dem gleichen Problem aussetzen, bis es gelöst ist. Wenn die Datenverarbeitung für ML vertraglich nicht gerechtfertigt werden könnte, kann es sein, dass SLU-Dienstleister nicht in der Lage wären, die vertraglichen Verpflichtungen zur Behebung von Mängeln in der den Kunden angebotenen Dienstleistung zu erfüllen. Dies wäre insbesondere dann äußerst problematisch, wenn die SLU-Genauigkeit an erster Stelle steht (z. B. im Gesundheitswesen).

Angesichts der Anzahl der Kunden, die denselben Dienst abonniert haben, muss ein bei der Erkennung eines von Kunde A ausgegebenen Sprachbefehls festgestellter Mangel daher auch für alle anderen Kunden genutzt werden können. Dies wird z. B. bei Software, die für alle Anwender parallel gepatcht wird, nicht diskutiert – die Situation bei ML ist genau die gleiche.

#### **4.5.5 Vertraglich vereinbarte kontinuierliche Verbesserung**

Das oben erläuterte Konzept zur Begründung der Vertragserfüllung gilt auch im Hinblick auf allgemeine Verbesserungen, da Kunden die Echtzeit-Lernprozesse als klassische Langzeit-Updates erwarten. Ein Kunde, der sich für ML-basierte Dienste anmeldet, meldet sich also nie für ein statisches, sondern für ein dynamisches Produkt oder eine Dienstleistung an, die sich z. B. im Hinblick auf neue Funktionalitäten ständig verbessert, die nur durch ein andauerndes Training von Algorithmen ermöglicht werden können. Die Kunden erwarten diese Verbesserung ebenso wie die Behebung von Mängeln. Da die vereinbarte Vertragserfüllung solche Verbesserungen beinhaltet, müssen auch hier die entsprechenden Kundendaten verarbeitet werden.

Auch hier gibt es eine Parallele zur Software-Entwicklung: Im Gegensatz zu greifbaren und statischen Produkten entwickeln sich die Funktionalitäten weiter und es ist oft schwer zu unterscheiden zwischen Fehlerbehebung, Wartung und Verbesserungen. Dies setzt jedoch unter Umständen natürlich auch die Verarbeitung der persönlichen Daten des Kunden voraus, um die Verbesserung überhaupt erst zu erreichen. Auch hier erfolgt die Verbesserung nicht nur gegenüber dem Kunden, der die Daten für das ML-Training bereitstellt, sondern gegenüber

allen teilnehmenden Kunden. Dies kann als eine Gruppenstruktur gesehen werden, bei der jedes Gruppenmitglied mehr Teil einer Gruppe als Individuum wird.

Das gleiche Prinzip des Gruppennutzens wird in Bereichen wie dem autonomen Fahren und der proaktiven Verkehrssteuerung angewendet. Diese funktionieren auch nur dann, wenn die von jedem Teilnehmer generierten Daten nicht nur gegenüber dem jeweiligen Teilnehmer verwendet werden, sondern auch, um Machine Learning zu generieren, das dann gegenüber den anderen Teilnehmern verwendet wird, z. B. um schädliche Situationen für alle zu vermeiden, wenn ein System ein potenziell schädliches Szenario erlernt hat.

ML kann daher als eine Form der vernetzten Dienstleistung verstanden werden, die allen Beteiligten zur Verfügung gestellt wird, aber auch den Input aller Beteiligten erfordert, um einen netzwerkspezifischen oder »Schwarm«-Nutzen zu erzielen.

Die Anbieter von ML-basierten Diensten verdanken ihren Kunden diesen zusätzlichen, auf Gruppendaten basierenden Nutzen. Er bildet eine Vertragserfüllungsanforderung und geht über die klassische Mängelbeseitigung hinaus, da es sich um einen dynamischen und kontinuierlichen Prozess handelt.

## 4.6 Gefahr der Restriktion durch die ePrivacy-Verordnung

Sofern in bestimmten Fällen Art. 6 Abs. 1 lit. b DS-GVO nicht greift, kann auch das berechtigte Interesse nach Art. 6 Abs. 1 lit. f DS-GVO und die kompatible Weiterverarbeitung gem. Art. 6 Abs. 4 DS-GVO einen Rahmen bieten, der eine entsprechende Datenverarbeitung ermöglichen kann. Je nach Rechtsgrundlage muss die Verarbeitung mit den ursprünglichen Zwecken vereinbar sein (Art. 6 Abs. 4 DS-GVO) oder sie muss einen Interessenausgleichstest bestehen (Art. 6 Abs. 1 lit. f DS-GVO). Der Entwurf der ePrivacy-Verordnung schließt diese Möglichkeiten aus und lässt ML nur in sehr engem Umfang zu, u. a. wenn dies vertraglich gefordert oder darin eingewilligt wurde. Da die ePrivacy-Verordnung die Rechtsgrundlagen der DS-GVO für die Verarbeitung von Daten auf allen angeschlossenen Kommunikationsgeräten, wie z. B. IoT, außer Kraft setzen könnte, ist die Investition in den Datenschutz, die auf der Grundlage des auf den DS-GVO-Grundsätzen beruhenden Vertrauens getätigt wurde, gefährdet.

Vor dem Hintergrund, dass die DS-GVO mit dem risikobasierten Ansatz einen Ausgleich zwischen den Interessen und Rechten der Nutzer und den legitimen Geschäftsinteressen der Anbieter herbeigeführt hat, ist es unverständlich, dass dies nun wieder stark beschränkt werden soll. Zur Unterstützung und im Vertrauen auf die von der DS-GVO geförderten Prinzipien haben viele Unternehmen auch durch gestalterische Maßnahmen wie die Pseudonymisierung stark in die Privatsphäre und den Datenschutz der Nutzer investiert.

Auf dieser Grundlage müssen Möglichkeiten der Weiterverarbeitung und der Schutz der Privatsphäre durch Designkonzepte auch für die ePrivacy-Verordnung übernommen werden.

Derzeit ist zwar noch nicht absehbar, wann die ePrivacy-Verordnung verabschiedet werden wird. Der ursprünglich von der EU-Kommission angedachte Zeitplan sah vor, dass die ePrivacy-Verordnung zeitgleich mit der DS-GVO ab 25.05.2018 zur Anwendung gelangen sollte. Zum jetzigen Zeitpunkt (Stand September 2018) hat jedoch der Rat der Europäischen Union (Ministerrat) noch keine Verhandlungsposition für den Trilog gefunden, die Mitgliedstaaten sind derzeit dabei, ihre Stellungnahmen zu erarbeiten bzw. fortzuentwickeln. Die Verhandlungen werden sich daher voraussichtlich noch einige Zeit hinziehen. Diese Zeit sollte genutzt werden, um die notwendigen Anpassungen vorzunehmen, um dann eine datenschutzgerechte aber zugleich innovationsoffene ePrivacy Verordnung zu verabschieden.

# 5 Privacy Enhancing Technologies

Verschiedene Technologien können eingesetzt werden, um die Arbeit mit personenbezogenen Daten auch im Zusammenhang mit Machine Learning-Verfahren sicherer zu machen, oder sie sogar ganz von weiteren Transparenzanforderungen zu befreien.

Hier sollen drei grundlegende Ansätze kurz vorgestellt werden:

- Pseudonymisierung
- Bisherige Ansätze zur Anonymisierung
- Neuartige Ansätze zur Anonymisierung

Aus Sicht der Transparenzanforderungen ergibt sich eine wichtige Unterscheidung: Erwägungsgrund 26 der DS-GVO bringt zum Ausdruck, dass pseudonymisierte Daten weiterhin als »Informationen über eine identifizierbare natürliche Person betrachtet werden« sollten. Dahingegen sollen die »Grundsätze des Datenschutzes [...] nicht für anonyme Informationen gelten«.

Dementsprechend bestehen die üblichen Transparenzanforderungen auch für pseudonymisierte Daten (dazu im folgenden Kapitel mehr); allerdings können Verantwortliche und Auftragsdatenverarbeiter dadurch eine höhere Sicherheit erzeugen und Missbrauch vorbeugen. Pseudonymisierung kann zudem als eine der in der DS-GVO zum Schutz personenbezogener Daten genannte notwendige technisch-organisatorische Maßnahme angewandt werden.

Anonymisierte Daten unterliegen hingegen keinen weiteren Anforderungen aus Datenschutzsicht. Eine korrekte Anonymisierung kann somit Verantwortliche und Auftragsverarbeiter extrem entlasten und eine weitere Verarbeitung – auch mit Machine Learning Algorithmen – datenschutzrechtlich uneingeschränkt ermöglichen. Allerdings ist sichere Anonymisierung bei gleichzeitigem Erhalt der Aussagekraft der Rohdaten fast unmöglich; schließlich ist es Ziel eines solchen Verfahrens, Details aus einem Datensatz zu entfernen.

Im nachfolgenden Abschnitt werden zwei verschiedene Varianten der Anonymisierung beschrieben: Eine "klassische" Anonymisierung mittels K-Anonymisierung, welche jedoch zum Teil als nicht ausreichend angesehen wird, und den neuartigen Ansatz zur Anonymisierung mittels Differential Privacy.

Es ist zu beachten, dass der Zeitpunkt der Anonymisierung entscheidend ist: Alle Verarbeitungsschritte vor einer solchen Verarbeitung sind weiterhin an die Datenschutzbestimmungen gebunden und es bedarf einer sensiblen Begutachtung der Rechtslage bezüglich der Verarbeitung der Daten zum Zwecke der Herstellung eines pseudonymisierten/anonymisierten Datensatzes.<sup>19</sup> Für die Verantwortlichen sind ein klares Verständnis der Architektur und eine korrekte Darstellung derselben daher entscheidend.

---

<sup>19</sup> Eine ausführliche Darstellung zu den datenschutzrechtlichen Implikationen, zu Pseudonymisierung und Anonymisierung findet sich im Bitkom Faktenpapier zu Blockchain und Datenschutz im Kapitel 3.1.2: <https://www.bitkom.org/noindex/Publikationen/2018/Leitfaeden/Blockchain-und-Datenschutz/180502-Faktenpapier-Blockchain-und-Datenschutz.pdf>.



## 5.1 Pseudonymisierung

Pseudonymisierung versucht Daten zu schützen, indem die Werte von direkten Identifikatoren (Personalausweisnummer, Name, etc.) eines Datensatzes durch Pseudonyme ersetzt werden. Diese Pseudonyme werden dabei über ein geeignetes Verfahren aus dem ursprünglichen Wert generiert oder komplett neu vergeben. Ein Pseudonym kann das gleiche Format (z. B. Erzeugung neuer Kunst-Namen oder neuer, gültiger Personalausweisnummern) wie der ursprüngliche Datentyp besitzen oder in einem anderen Format vorliegen (z. B. zufällige Zeichenabfolgen). Wichtig ist lediglich, dass die Zuordnung eindeutig ist, dass also für zwei identische Eingabewerte immer das gleiche Pseudonym erzeugt wird. Für viele Anwendungen muss eine Pseudonymisierung zusätzlich umkehrbar sein, d. h. es muss möglich sein aus dem Pseudonym, gegebenenfalls mit zusätzlichen Informationen wie einem Schlüssel oder einer Tabelle, wieder den ursprünglichen Datenwert abzuleiten.

Ein Beispiel: Um Patienten im Rahmen einer medizinischen Studie zu schützen werden die Namen der Patienten pseudonymisiert, d. h. aus Max Mustermann wird beispielsweise P14924. Die an der Studie beteiligten Forscher, die Datensätze zu Patient P14924 auswerten, können dann sicher sein, dass diese alle zur gleichen Person gehören, sie können jedoch nicht ableiten, dass es sich dabei um Max Mustermann handelt.

Das Diagramm zeigt zwei Tabellen, die durch einen Pfeil verbunden sind. Der Pfeil ist beschriftet mit 'Ersetzen von PIIs<sup>20</sup> durch andere eindeutige Identifier'. Die linke Tabelle hat die Spalten 'Name' und '...' und enthält die Namen Max Mustermann, Petra Panther, Ali Anker, Brad Bär und Erna Edelstein. Die rechte Tabelle hat die Spalten 'Name' und '...' und enthält die Identifikatoren P14924, P14925, P14926, P14927 und P14928.

Name	...
Max Mustermann	
Petra Panther	
Ali Anker	
Brad Bär	
Erna Edelstein	

Ersetzen von PIIs<sup>20</sup> durch andere eindeutige Identifier

Name	...
P14924	
P14925	
P14926	
P14927	
P14928	

Abbildung 8: Beispiele Pseudonymisierung, adaptiert von Aircloak

<sup>20</sup> Personally Identifiable Information.

Pseudonymisierung wird also vorwiegend eingesetzt um sensitive Daten bei der Verarbeitung vor direkten Blicken zu schützen: Pseudonymisierung macht es schwerer, Rückschlüsse auf einen ursprünglichen Datenwert zu ziehen, bewahrt aber die Eindeutigkeit dieses Wertes und erhält so (teilweise) die Nutzbarkeit der Daten. Pseudonymisierung schützt im Gegensatz zu Anonymisierung jedoch nicht vor statistischen Angriffen zur Re-Identifikation von einzelnen Datenwerten. Wüsste in unserem Beispiel z. B. ein Arzt, dass Max Mustermann genau 1,80 m groß, 34 Jahre alt und 78 kg schwer ist, und trifft dies in den pseudonymisierten Daten nur auf Patient P14924 zu, so könnte er daraus ableiten, dass sich Max Mustermann hinter diesem Pseudonym verbirgt. Die Kombination anderer Attribute eines Datensatzes (z. B. Alter, Geschlecht und Postleitzahl), welche für sich allein nicht zur Identifizierung geeignet sind, kann einen sogenannten Quasi-Identifikator erzeugen, der durch die Kombination mehrerer Attribute dann ein (beinahe) eindeutiger Identifikator wird.

Pseudonymisierte Daten haben daher aus Gründen der Re-Identifizierbarkeit immer noch einen Personenbezug und unterliegen im Gegensatz zu anonymisierten Daten weiterhin der DS-GVO und weiteren datenschutzrechtlichen Vorschriften.

Die Pseudonymisierung ist aus datenschutzrechtlicher Sicht aber ein sehr wichtiges Verfahren. Die DS-GVO erwähnt Pseudonymisierung in den Artikeln 25, 26 und 40 sowie in Erwägungsgrund 28 und hebt die Wichtigkeit dieses Verfahrens bei der Verarbeitung personenbezogener Daten hervor. Insbesondere soll Pseudonymisierung als technische Maßnahme zur Sicherung von Daten eingesetzt werden um die Rechte betroffener Personen zu schützen.

Zusammengefasst sind die Vorteile von Pseudonymisierung, dass pseudonymisierte Daten sehr viel besser vor Verlust und Missbrauch geschützt sind als die Ursprungsdaten und dementsprechend auch einfacher und mit weniger Bedenken genutzt werden können, als die Klardaten. Pseudonymisierung erhält im Gegensatz zu Anonymisierung einzelne Datenpunkte (z. B. die Daten einzelner Personen) und fasst diese nicht mit anderen zusammen. Für viele Anwendungen können pseudonymisierte Daten daher genau wie die Ursprungsdaten verwendet werden. Nachteil von Pseudonymisierung gegenüber Anonymisierung ist, dass die Daten weiterhin einen Bezug zum ursprünglichen Datenwert (z. B. der Person zu der die Daten gehören) aufweisen, was auch die weitere Anwendbarkeit der DS-GVO begründet. Pseudonymisierung schützt Daten im Gegensatz zu Anonymisierung zudem nicht vor einer Re-Identifikation mithilfe statistischer Angriffe.

Vorteile	Nachteile
Pseudonymisierte Daten sind gut vor Verlust und Missbrauch geschützt	Pseudonymisierte Daten sind weiterhin personenbezogen
Pseudonymisierung erhält teilweise die Nutzbarkeit der Daten	Kein Schutz gegenüber statistischen Angriffen

Tabelle 1: Vor- und Nachteile von Pseudonymisierung

## 5.2 K-Anonymisierung

K-Anonymität versucht das Problem der Quasi-Identifikatoren (siehe vorheriges Kapitel zur Pseudonymisierung) zu umgehen. K-Anonymität ist eine Eigenschaft einer Datenbank die einen gewissen Grad der Anonymisierung darstellt. Das Modell besagt, dass jeder Eintrag über eine Person in der Datenbank nicht von K-1 anderen zu unterscheiden sein sollte.<sup>21</sup> Die Idee ist folgende: wenn Daten derart zusammengefasst sind, dass K Personen gleich aussehen, ist die individuelle Privatsphäre jedes Datensubjekts geschützt. In der K-Anonymität ist der Wert von K die Maßeinheit für Anonymität. Wenn  $K = 1$ , dann sind die Originaldaten unverändert und es gibt keine Anonymität – je größer K, desto besser ist die Anonymität.

Dass einzelne Personen durch die Kombinationen von Attributen zu identifizieren sind, wurde bereits im vorherigen Kapitel erwähnt. Ein konkretes Beispiel beschreiben Xiao<sup>22</sup> und Golle<sup>23</sup>: So sind 63 % aller Amerikaner anhand ihres Alters, Geschlechtes und Geburtsortes eindeutig identifizierbar. Man versucht daher diese Kategorien so zusammenzufassen, dass sie nicht mehr auf Einzelpersonen zurückführen.

---

21 Latanya Sweeney. 2002. »k-anonymity: A model for protecting privacy«, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10, 05 (2002), 557–570.

22 Xiaokui Xiao, Privacy Preserving Data Publishing: From k-Anonymity to Differential Privacy, abrufbar unter: <https://pdfs.semanticscholar.org/presentation/b735/c888d7f389744ca8644fb81b89372eba676e.pdf>.

23 Philippe Golle, Revisiting the uniqueness of simple demographics in the US population, veröffentlicht in: Proceeding WPES '06 Proceedings of the 5th ACM workshop on Privacy in electronic society, 77–80.

Das Verfahren sei im folgenden Beispiel erklärt:

Birthdate	Zip	Salary	
1979-11-17	14171	€ 15811	
1992-07-14	13925	€ 174731	→
1998-12-28	13926	€ 48291	
1982-01-31	28635	€ 84491	

Ausschnitt aus der originalen, nicht anonymisierten Tabelle A

Birthdate	Zip	Salary	
1998-12	13***	€ 30–40 K	(4 users)
1998-12	13***	€ 40–50 K	(7 users)
1998-12	13***	€ 50–60 K	(4 users)
1998-12	13***	€ 60–70 K	(6 users)

K-anonymer Ausschnitt aus der originalen Tabelle A

Birthdate	Zip	Salary	
1998-11	13***	€ 30–40 K	(4 users)
1998-12	13***	€ 30–40 K	(7 users)
1998-12	13***	€ 80–90 K	(4 users)
1998-13	13***	€ 70–80 K	(6 users)

K-anonymer Ausschnitt aus der originalen Tabelle B

Abbildung 9: Beispieltabelle K-Anonymisierung. Die Abbildung veranschaulicht diesen Ansatz. Oben links ist ein Ausschnitt aus einer originalen Tabelle A mit drei Spalten (Geburtsdatum, Postleitzahl und Gehalt). Oben rechts ist ein Ausschnitt der K-anonymisierten Tabelle A. Im unteren Teil der Abbildung sind Ausschnitte aus zwei weiteren K-anonymisierten Tabellen aus anderen Originaltabellen B zu sehen. Adaptiert von Aircloak

Angenommen der Analyst Andreas weiß, dass seine Kollegin Sandra am 28.12.1998 Geburtstag hat und unter der Postleitzahl 13926 wohnt, er aber nicht ihr Gehalt kennt. Außerdem gelangt er in Besitz der Tabelle A aus Abbildung 9. Man beachte, dass Tabelle A keine direkten Identifikatoren wie Namen, Personalnummern oder Adressen enthält. Trotzdem erfährt Andreas nun genau das Gehalt von Sandra: € 48.921. Die Kombination aus Postleitzahl und Geburtsdatum fungierten hier als Quasi-Identifikatoren.

In der K-anonymisierten Tabelle A (rechts) wird versucht, das Risiko der Quasi-Identifikation und der Zusammenführung zu eliminieren. Hier existieren nun viele Personen, die im Dezember 1998 geboren wurden und in der zweistelligen Postleitzahl 13\*\*\* wohnen. Außerdem existieren mehrere unterschiedliche Gehaltsgruppen für diese Kombinationen. Andreas kann so nicht mehr auf das Gehalt von Sandra rückschließen. Hier wurden die Werte für eine k-anonymisierte Tabelle erfolgreich generalisiert und gruppiert, sodass Daten geschützt wurden. In der K-anonymisierten Tabelle A sind mindestens vier Personen zu einer Kombination zugehörig, wodurch hier eine 4-Anonymität vorliegt.

### 5.2.1 Kritik

Die Kritik an K-Anonymität lässt sich gut am laufendem Beispiel basierend auf Abbildung 9 erläutern: In Tabelle B (ebenfalls 4-Anonym) fallen alle im Dezember 1998 geborenen Personen und Postleitzahl 13\*\*\* in zwei Gehaltsbereiche, wodurch nach wie vor eine K-anonymisierte Tabelle vorliegt. Hier ist es Andreas nun trotzdem möglich Sandras Gehalt bis auf € 10.000 einzugrenzen: Es existieren zwar zwei mögliche Gehaltsbereiche (30–40 k€ und 80–90 k€), Andreas weiß jedoch, dass Sandra eine Programmiererin ist, deren Gehälter mindestens € 50 K betragen müssen laut Tarifvertrag. Durch dieses Hintergrundwissen, kann er ableiten, dass Sandras Gehalt zwischen € 80 K und € 90 K liegen muss. Aus diesen Beispielen wird sichtbar, dass K-Anonymität nicht immer den Rückschluss auf persönliche Daten verhindert.

Das Problem von Hintergrundwissen wird dadurch verstärkt, dass wenn viele Spalten existieren, die Anzahl der möglichen Kombinationen exponentiell steigt (wenn es nur 10 Spalten/Attribute gibt, mit jeweils nur 5 Antwortmöglichkeiten als Wert, dann existieren bereits  $5^{10} = 9.765.625$  verschiedene Kombinationen). Für jede dieser auftretenden Kombinationen müssen mindestens K (z. B. 4) Personen vorliegen, da sonst keine K-Anonymität mehr gewährleistet werden kann! Die so notwendige starke Aggregation führt oft dazu, dass jeglicher Informationsgehalt der Tabellen verloren geht. Manchmal wird so versucht K-Anonymität nur auf Spalten, welche Quasi-Identifikatoren darstellen könnten, anzuwenden. Dies setzt voraus, dass man Spalten in »potentiell bekannt« und »potentiell unbekannt und schützenswert« einteilen kann. Das Problem von Hintergrundwissen ist jedoch, dass dadurch theoretisch jede Spalte »potentiell bekannt« wird, wodurch sich wiederum vorheriges Problem ergibt.

### 5.2.2 K-Anonymisierung und Machine Learning

In einer K-anonymisierten Tabelle existiert für jede Zeile der Originaltabelle eine Zeile in der anonymisierten Datenbank. Daher kann jede Analyseverfahren, die auf der Originaltabelle laufen könnte, prinzipiell auch auf der K-anonymisierten Tabelle funktionieren – auch Machine Learning Ansätze. Wenn die K-Anonymisierung allerdings die Aussagekraft der Daten wesentlich verringert hat, leidet das Ergebnis von ML mindestens in gleichem Maße.

Vorteile	Nachteile
Einfach verständliches Konzept	In der Praxis komplex zu konfigurieren für Aggregationsverfahren
Kann genutzt werden, um pseudonymisierte Daten sicherer zu machen	Führt oft zu sehr schlechter Datenqualität, wenn sämtliche Spalten anonymisiert werden
Kann gute aggregierte Resultate geben	Durch Hintergrundwissen angreifbar
(Open Source-)Software verfügbar	

Tabelle 2: Vor- und Nachteile der K-Anonymität

Die Nachteile der K-Anonymität werden durch den radikal anderen Ansatz der Differential Privacy angegangen. Dies soll im Folgenden Kapitel erläutert werden.

## 5.3 Differential Privacy

Der Kerngedanke hinter Differential Privacy ist, jegliche Zugehörigkeit zu einem Datensatz oder zu einem Ergebnis plausibel abstreiten zu können. Man versucht so nicht zu verhindern, dass eine Datenzeile auf eine Person hinweist, was bereits häufig versucht wurde, jedoch immer fehlschlug, sondern schafft die Möglichkeit (lies: Wahrscheinlichkeit), dass eine Datenzeile gar nicht echt ist. Genauer: Die Wahrscheinlichkeit, dass eine Datenzeile oder ein Ergebnis auftaucht ist nahezu unabhängig davon, ob die Daten einer Person überhaupt in der Datenbank eingefügt wurden. Mathematisch liest sich dies wie folgt:

### Definition Differential Privacy Algorithmus

Sei  $\vec{X}$  eine Datenbank an Datenzeilen  $\vec{x}$ . Sei  $\vec{X}^*$  eine Datenbank in der eine einzige Zeile  $\vec{x}^*$  unterschiedlich zu  $\vec{X}$  sei. Sei  $r$  ein Analyseergebnis, welches aus der Datenbank  $\vec{X}$  oder  $\vec{X}^*$  ziehbar ist. (z. B. eine Abfrage einer einzelnen Datenzeile oder die Summe einer Spalte).

Dann ist ein Algorithmus »differentially private« wenn gilt:

$$P(r|\vec{X}) \leq P(r|\vec{X}^*) \cdot e^\epsilon$$

Anmerkung: Für Abfragen auf einzelne Datenzeilen reduziert sich die bedingte Wahrscheinlichkeit von  $\vec{X}$  auf  $\vec{x}$ .<sup>18</sup>

Es empfiehlt sich hier nochmals den zentralen Kerngedanken dieses Theorems zu wiederholen: Aus einem Datensatz der mittels Differential Privacy geschützt wurde bzw. wird, ist jegliches Ergebnis nahezu gleich wahrscheinlich. Erfährt man beispielsweise, dass Datenzeile 2834 einer geschützten Datenbank zu Max Mustermann gehört, weil aus irgendwelchen Gründen eine Ursprungsperson-Zeilen-Beziehung leaked wurde (z. B. durch Hintergrundwissen über das Alter in Abbildung 10), besteht für Max kein akutes Problem: seine Datenzeile wurde so verfremdet (siehe Abbildung 10, Daten wurden durch DP verändert-rot), dass die Möglichkeit besteht, dass sie zufällig generiert wurde. So kann er eine hypothetische geleakte Information über sein Gehalt, mögliche Krankheiten, Wohnort, Alter, etc. glaubhaft abstreiten (siehe Abbildung 10, unten rechts), da eine signifikante Wahrscheinlichkeit existiert, dass diese zufällig eingesetzt und verfremdet wurde.

Aus selbigem Argument fallen auch Verknüpfungsangriffe (wie zuvor bei K-Anonymität) durchs Raster: Die Kombination aus Alter, Geschlecht und Wohnort ist nach wie vor identifizierend, aber

<sup>18</sup> Jane Bambauer, Krishnamurty Muralidhar, Rathindra Sarathy, Fool's Gold: An Illustrated Critique of Differential Privacy, abrufbar unter: [http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer\\_Final.pdf](http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer_Final.pdf); Cynthia Dwork, Aaron Roth, The Algorithmic Foundations of Differential Privacy, abrufbar unter: <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>; Frank McSherry, Kunal Talwar, Mechanism Design via Differential Privacy, abrufbar unter: <http://kunalatalwar.org/papers/expmech.pdf>.

jede Kombination in der Datenbank kann verfälscht sein. Das charmante ist jedoch: Über die gesamte Datenbank hinweg bleiben Statistiken erhalten (siehe Abbildung 10, Verteilungen im Ergebnis mit (blau) und ohne (hellgrau) sind nahezu identisch). Detaillierte Analysen über die gesamte Bevölkerung – und nicht auf Max speziell – bleiben genau und aussagekräftig. Außerdem können in vielen Fällen bestehende Analyseverfahren direkt weiterverwendet werden (siehe Abbildung 10, das identische Analyseverfahren wird zur Auswertung verwendet). Darüber hinaus können mittels auf DP angepassten Verfahren Ergebnisse erzielt werden, die nahezu identisch sind zu den originalen Ergebnissen (siehe Abbildung 10, »DP angepasste Analyse«).

Zusammengefasst, wie in Abbildung 10 zu sehen, werden durch einen Differential Privacy Algorithmus Daten derart verfremdet, dass eine einzelne Zeile (Person) für sich genommen wertlos wird, Aussagen über alle Zeilen zusammen jedoch dasselbe (oder ein sehr ähnliches) Ergebnis liefern wie ohne eine DP-Anonymisierung; wie diese Genauigkeit trotz Verfremdung der Daten erreicht wird, dazu mehr nach folgendem Exkurs.

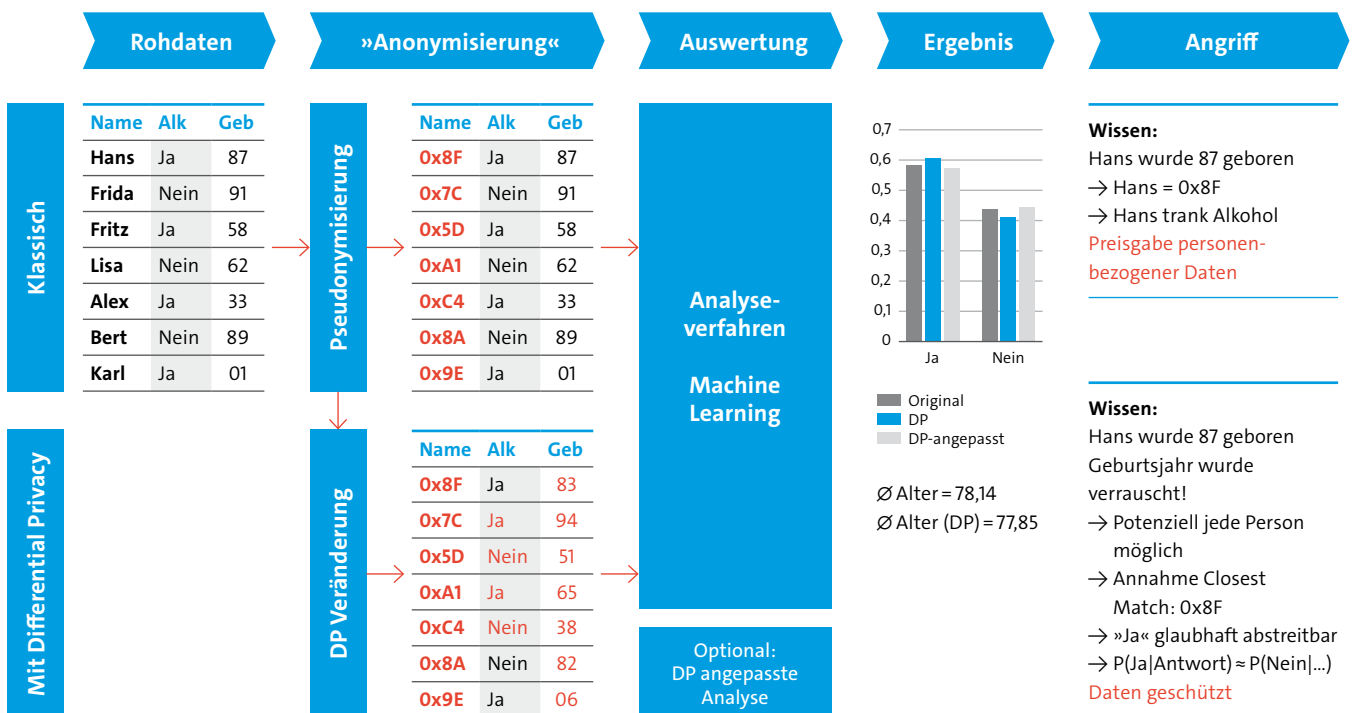


Abbildung 10: Datenfluss mit und ohne Differential Privacy Ansätze. Nach einer Befragung von Personen ob sie generell Alkohol konsumieren, werden Analysen aus den Umfrageergebnissen gewonnen. Wird der Datensatz nur pseudonymisiert veröffentlicht, können nach wie vor einzelne Personen genau anhand ihres Alters identifiziert werden (ähnliche Angriffe gelten auch für K-Anonymität). Mittels Differential Privacy können solche Hintergrundwissensangriffe verhindert werden, Analyseergebnisse bleiben nahezu identisch und viele bestehende Analysen können ungehindert weiter durchgeführt werden. Adaptiert von Lufthansa Industry Solutions

### 5.3.1 Differential Privacy ist kein Algorithmus

Ein wichtiger Punkt der DP-Definition ist, dass Differential Privacy eine mathematische Eigenschaft eines Algorithmus ist. Sie ist keine Eigenschaft einer Datenbank wie bei K-Anonymität oder Pseudonymisierung, sondern qualifiziert ein Verfahren. Außerdem hat diese Eigenschaft verschiedene Ausprägungen ( $\epsilon$ ). Es ist somit nicht möglich »Differential Privacy« anzuwenden; es muss gezielt ein spezieller Algorithmus ausgewählt werden und dieser entsprechend der gewünschten Ausprägung eingestellt werden.

Der zentrale Vorteil von etablierten Differential Privacy Algorithmen ist, dass sie die Daten oder Ergebnisse nur so weit verfremden, dass Forscher noch immer gute Ergebnisse aus den Daten ziehen können.

An zentraler Stelle stehen hier zwei Richtungen bzw. Sichtweisen auf Differential Privacy Algorithmen für die Anonymisierung von Daten oder Ergebnissen:

1. Interaktive Verfahren zur Anonymisierung von Ergebnissen und
2. Statische Verfahren zur Anonymisierung von Datensätzen

Mathematisch betrachtet sind beide Fälle identisch: Beide Fälle sind Anfragen eine Datenbank. In einem Falle (1) wird dabei eine einzelne Datenzeile und im anderen (2) ein Ergebnis über mehrere Zeilen hinweg abgefragt. In der Literatur werden beide Fälle jedoch meist gesondert betrachtet.<sup>25</sup>

Aus technologischer Sicht empfiehlt sich eine getrennte Betrachtung ebenfalls, da

1. einen »Proxy« bedarf und ständig wiederholt durchgeführt werden muss,
2. ein einmaliger Vorgang ist.

Die interaktive Anonymisierung ist grundsätzlich verschieden zu klassischen Ansätzen wie K-Anonymität oder Pseudonymisierung. Bei der interaktiven Anonymisierung liegen in einer stark bewachten Datenbank die echten, unanonymisierten Daten nach wie vor vor. Möchten Forscher nun eine Analyse auf dieser Datenbank durchführen, wird diese Analyse auf den Echtdateien durchgeführt. Es findet also eine Verarbeitung von personenbezogenen Daten statt und das Ergebnis wird mittels eines DP-Algorithmus so verändert, dass das Ergebnis (z. B. ein Histogramm, Mittelwert, Summe, Korrelationsmatrix, etc.) keine persönlichen Informationen preisgibt. Dies ist ein extrem wertvolles Verfahren um Analyseergebnisse auszutauschen ohne hierbei selbst personenbezogene Informationen zu »leaken«. Da dies jedoch ein hoch komplexes

---

<sup>25</sup> Cynthia Dwork, A firm foundation for private data analysis, veröffentlicht in: Magazine Communications of the ACM Volume 54, Issue 1, January 2011 Pages 86–95; Rathindra Sarathya and Krishnamurty Muralidhar, Evaluating Laplace Noise Addition to Satisfy (Differential Privacy for Numeric Data), abrufbar unter: <http://www.tdp.cat/issues11/tdp.a064a10.pdf>; Jane Bambauer, Krishnamurty Muralidhar, Rathindra Sarathy, Fool's Gold: An Illustrated Critique of Differential Privacy, abrufbar unter: [http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer\\_Final.pdf](http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer_Final.pdf); Yue Wang, Xintao Wu, Donghui Hu, Using Randomized Response for Differential Privacy Preserving Data Collection, abrufbar unter: <http://ceur-ws.org/Vol-1558/paper35.pdf>.



System erfordert und darüber hinaus personenbezogene Daten verarbeitet werden, betrachten wir im Weiteren nur den zweiten Fall der statischen Anonymisierung.

### 5.3.2 No Free Lunch

Die statische Anonymisierung ist eine klassische Anonymisierung, bei der ein neuer Datensatz erzeugt wird, in welchem jeglicher Personenbezug entfernt worden ist und so als anonym angesehen wird; jedenfalls nach den besten Kenntnisstand der Wissenschaft und den Ausprägungen des verwandten DP-Verfahrens. Man erhält also eine neue Datenbank, welche frei verteilbar wäre und aus welcher nach wie vor beste Analyseergebnisse gezogen werden können. Beispiele für adäquate Verfahren sind das Randomized-Response-Verfahren oder der Laplace-Mechanismus. Ersterer ist sehr einfach und intuitiv zu erklären und stammt aus der Sozialwissenschaft um Interviews mit potentiell unangenehmen Antworten durchzuführen.<sup>26</sup> Ein Befragter soll eine Frage beantworten, wohingegen die Antwort für ihn jedoch beschämend sein kann, z. B. ob man zum Beispiel eine bestimmte Partei wähle. Personen die eigentlich beabsichtigen diese Partei zu wählen sind unter Umständen dazu geneigt eine Falschantwort zu geben. Dem Befragten wird mittels des Randomized-Response-Verfahren nun die Möglichkeit gegeben zu »Lügen«, jedoch unter Laborbedingungen. Der Befragte wirft zwei Mal eine Münze. Erscheint beim ersten Wurf Kopf, so antwortet er die Wahrheit, wirft er jedoch Zahl so wird er eine zufällige Antwort geben basierend auf dem Ergebnis des zweiten Wurfs. Egal was er antwortet, er behält zu jedem Zeitpunkt die Möglichkeit, es auf ein »Das war nur die Münze« abzuspielen. Es zeigt sich, dass genau dieses Verfahren  $\epsilon = \ln(3)$  Differential Privacy erfüllt. Dadurch, dass das Lügen unter Laborbedingungen ausgeführt wird, d. h. dass man die Wahrscheinlichkeiten des Lügenprozesses kennt, lässt sich dieser jedoch aus allen Antworten herausrechnen: Wenn nun  $p'$  % der befragten als potentielle Wähler antworteten, ergibt sich die angenäherte echte Verteilung  $p$  zu  $p = p' \cdot 2 - 0.5$ .<sup>27</sup>

Der Teufel steckt jedoch im Detail. Im vorherigen Beispiel wurde eine  $\epsilon = \frac{2}{3}$  DP durch die verwendete Münze erreicht. Es ist leicht vorstellbar, dass hier auch ein  $n$ -seitiger Würfel für den zweiten Wurf verwendet wird um  $n$  Antwortmöglichkeiten zu repräsentieren. Dieser Würfel kann ebenfalls gezinkt sein. Mit diesen beiden Stellschrauben ist theoretisch jeder  $\epsilon$ -Wert erreichbar.

Es stellt sich daher die zentrale Frage welcher  $\epsilon$ -Wert ausreichend ist um von einer sauberen Anonymisierung zu sprechen. Bei einem  $\epsilon = \infty$  ergibt sich absolut keine Anonymisierung, bei  $\epsilon = 0$  werden die Daten vollständig vernichtet, bzw. mit vollständig informationsvernichtenden Daten überschrieben. Die Wahrheit liegt somit irgendwo zwischen diesen beiden Welten und bedarf eines sauberen Angreifermodells mit entsprechenden Angriffsverfahren um die Wirksamkeit der  $\epsilon$ -DP-Anonymisierung nachweisen zu können. Bei kontinuierlichen

<sup>26</sup> Stanley L. Warner, Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias, Journal of the American Statistical Association Vol. 60, No. 309 (Mar., 1965), 63–69.

<sup>27</sup> Cynthia Dwork, Aaron Roth, The Algorithmic Foundations of Differential Privacy, abrufbar unter: <https://www.cis.upenn.edu/~aaroht/Papers/privacybook.pdf>.

Wertebereichen wird oftmals auf das sogenannte Laplace-Verfahren zurückgegriffen, welches einen kontinuierlichen Zahlenwert, gezogen aus einer Wahrscheinlichkeitsverteilung, auf den echten Wert hinzuaddiert. Die notwendige Wahrscheinlichkeitsverteilung für eine  $\epsilon$  differentielle Privatheit bestimmt sich durch den möglichen Wertebereich einer Variable. Dies stellt ein zentrales Problem dar, denn das Gehaltsspektrum aller Personen kann nicht allgemeingültig bestimmt werden: Es wäre ein so riesiger Bereich, dass »normale« Gehälter plötzlich zu »anonymen« plus/minus mehreren Millionen Euro würden. Einzelne Personen auszuschließen um den Wertebereich einzudämmen ist jedoch auch keine Lösung: Wenn bekannt würde, dass eine Person nun nicht Teil der Datenbank ist, dies aber sein sollte, so ist klar, dass sein Gehalt die maximale Aufnahmeschwelle in die Datenbank überschritten hat.

→ Auch durch das Hinweglassen von Informationen können Informationen publik werden.

Besondere Sorgfalt ist außerdem bei korrelierten Daten notwendig. Sind einzelne Zeilen miteinander verbandelt (korreliert), z. B., da die Befragten alle einer Clique oder eines sozialen Netzwerkes entstammen, können einzelne Zeilen nicht mehr für sich selbst anonymisiert werden. Selbiges gilt für korrelierte Spalten (z. B. Größe und Gewicht). Bei solch verbandelten Spalten oder Zeilen sind manche Kombinationen deutlich wahrscheinlicher als andere. Werden Attribute einzeln anonymisiert, so liefert das Vorwissen über wahrscheinlichere und unwahrscheinlichere Attributs- oder Zeilenkombinationen Wissen darüber was die Ursprungsdaten gewesen sein könnten.

Während in der Literatur und hier viel über Verfahren und Algorithmen zur Differential Privacy gesprochen wird, so bleibt es nach wie vor ein mathematisches Konstrukt. Ein angedachtes Verfahren wird in einen Wahrscheinlichkeitsprozess zerlegt, für welchen ein mathematischer Beweis »auf Papier« geführt wird.

Jeder dieser Abstraktionsschritte birgt Fehlerrisiken, denn wenn die letztendliche Implementierung des Computeralgorithmus nicht haargenau dem ursprünglichen mathematischen Modell entspricht, so ist auch der Beweis und dessen Anonymisierungsfähigkeit hinfällig. Durch die Deterministik bestehender Computer sind vor Allem solch probabilistische Verfahren mit äußerster Sorgfalt zu implementieren. Besonders beeindruckende Fehlschläge für gescheiterte Implementierung von Differential Privacy liefern z. B. Bambauer, Muralidhar und Sarathy.<sup>28</sup> Die Liste an bisher gescheiterten Anonymisierungsversuchen mittels anderen Verfahren sollte jedoch auch nicht unterschätzt werden.<sup>29</sup>

---

28 Jane Bambauer, Krishnamurty Muralidhar, Rathindra Sarathy, Fool's Gold: An Illustrated Critique of Differential Privacy, abrufbar unter: [http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer\\_Final.pdf](http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer_Final.pdf).

29 Xiaokui Xiao, Privacy Preserving Data Publishing: From k-Anonymity to Differential Privacy, abrufbar unter: <https://pdfs.semanticscholar.org/presentation/b735/c888d7f389744ca8644fb81b89372eba676e.pdf>; Philippe Golle, Revisiting the uniqueness of simple demographics in the US population, veröffentlicht in: Proceeding WPES '06 Proceedings of the 5th ACM workshop on Privacy in electronic society, 77–80.

Durch die vorhergehenden Punkte wird deutlich, dass der nachweisbare Einsatz von »Differential Privacy« mit Aussagen über  $\epsilon$ , das verwendeten Verfahren, zugehörigen Beweisen und der Offenlegung der Implementierungen verbunden werden sollte.<sup>30</sup>

Wie oben dargelegt, ist eine Implementierung von Differential Privacy, die sowohl starke Sicherheit als auch eine gute Datenqualität gewährleistet, eine große Herausforderung.

### 5.3.3 Zusammenfassung


Differential Privacy ist eine mathematische Definition, die für einen Algorithmus erfüllt sein kann. Wendet man solch einen Algorithmus auf einen Datensatz an, so führt dies zu einer Anonymisierung von Daten, die dem Stand der Technik entspricht und bisher jeglichen Angriffen widerstanden hat, solange die Rahmenbedingungen und Annahmen der Differential Privacy Definition befolgt werden. Die Praxis<sup>31</sup> zeigt, dass diese Annahmen und Rahmenbedingungen schnell durch kleine Fehler verletzt werden können oder schlecht angepasste Parameter ( $\epsilon$ ) gewählt werden, wodurch entweder die anonymisierten Daten wertlos werden für Datenanalysen oder keine genügende Anonymisierung, d. h., ein Schutz von personenbezogenen Daten, mehr existiert.

Gelingt die Gratwanderung adäquate Parameter ( $\epsilon$ ) zu wählen und werden alle Annahmen, Rahmenbedingungen, Implementierungen und Beweise sauber eingehalten und durchgeführt, so schafft die Differential Privacy den heiligen Gral des Datenschutzes und Daten Analysen: Personenbezogene Daten werden geschützt unter dem Erhalt detaillierter Analysemöglichkeiten und Einblicken in die Daten.

---

30 Privacy at the End of the Rainbow: <https://iapp.org/news/a/differential-privacy-at-the-end-of-the-rainbow/> – Francis, Paul.

31 Jane Bambauer, Krishnamurty Muralidhar, Rathindra Sarathy, Fool's Gold: An Illustrated Critique of Differential Privacy, abrufbar unter: [http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer\\_Final.pdf](http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer_Final.pdf).



Bitkom vertritt mehr als 2.600 Unternehmen der digitalen Wirtschaft, davon gut 1.800 Direktmitglieder. Sie erzielen allein mit IT- und Telekommunikationsleistungen jährlich Umsätze von 190 Milliarden Euro, darunter Exporte in Höhe von 50 Milliarden Euro. Die Bitkom-Mitglieder beschäftigen in Deutschland mehr als 2 Millionen Mitarbeiterinnen und Mitarbeiter. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 400 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig oder in anderer Weise Teil der digitalen Wirtschaft. 80 Prozent der Unternehmen haben ihren Hauptsitz in Deutschland, jeweils 8 Prozent kommen aus Europa und den USA, 4 Prozent aus anderen Regionen. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem weltweit führenden Digitalstandort zu machen.

**Bundesverband Informationswirtschaft,  
Telekommunikation und neue Medien e.V.**

Albrechtstraße 10  
10117 Berlin  
T 030 27576-0  
F 030 27576-400  
bitkom@bitkom.org  
[www.bitkom.org](http://www.bitkom.org)

**bitkom**