Künstliche Intelligenz und Datenschutz

Martin Rost
(mit Folien auch von Benjamin Walczak und Dr. Probst)

Lübeck, den 09.11.2018

„Datenschutz hat in 30 Jahren normativer und operativer Aktivitäten die Anforderungen ausgebildet, um Systeme der künstlichen Intelligenz beherrschbar zu machen.“

Agenda

1. Einführung: Was meint “Datenschutz”?
2. Was meint “Künstliche Intelligenz” (KI) oder “Maschinenlernen” (ML)?
3. Beispiele für intelligente Systeme
4. Wie funktionieren intelligente Systeme?
5. Anforderungen des Datenschutzes an KI-Systeme
6. Versprengtes
7. Referenzen

Agenda

1. Einführung: Was meint “Datenschutz”?
2. Beispiele für intelligente Systeme
3. Wie funktionieren intelligente Systeme?
4. Was ist aus Datenschutzsicht das Problem bei KI/ML?
5. Versprengte
6. Referenzen

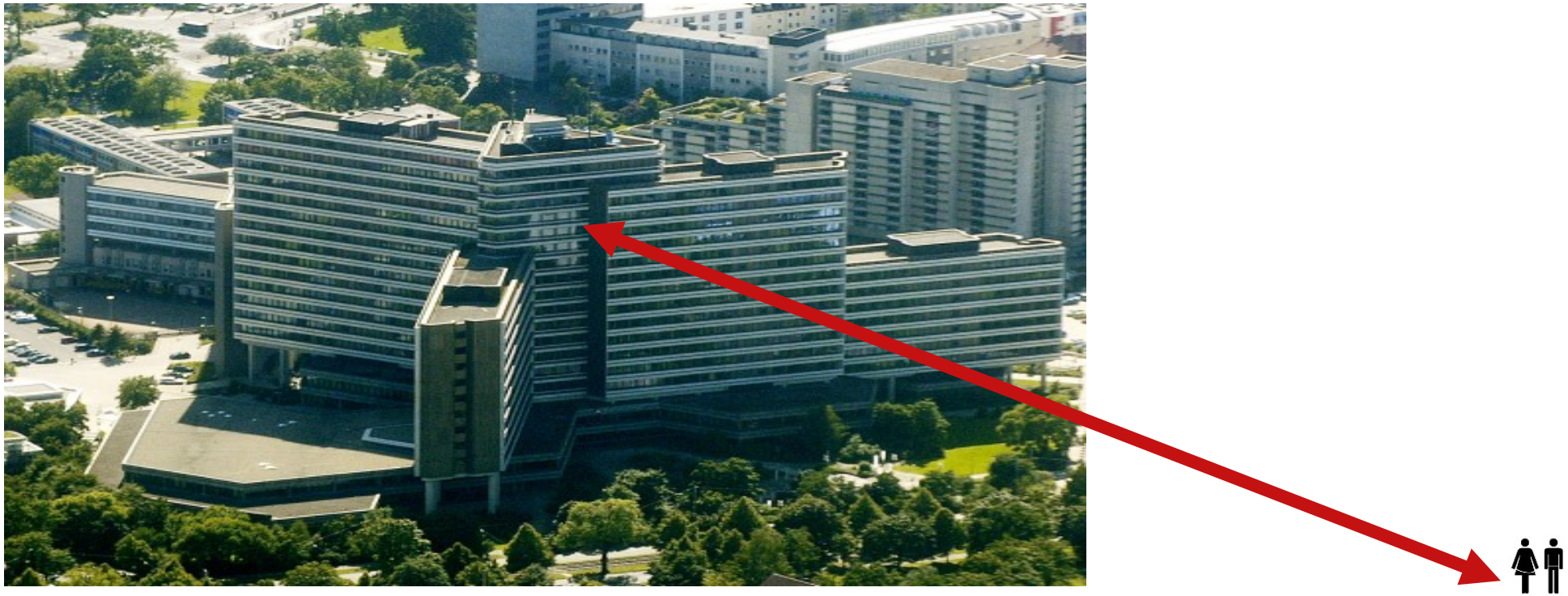
Was meint „Datenschutz“?

- **Datenschutz ist nicht mit Datenschutzrecht gleichzusetzen!**
  Denn das Datenschutzrecht reagiert auf einen strukturellen Konflikt. Nur: Worin genau besteht dieser strukturelle Konflikt des Datenschutzes, den das Datenschutz-Recht lösen soll?

- **Datenschutz ist nicht mit der IT-Sicherheit gleichzusetzen!**
  Ein technisch sehr gut abgesichertes System kann vollkommen unrechtmäßig betrieben werden. Deshalb ist mit Konflikten zw. Datenschutz und IT-Sicherheit zu rechnen. Rechtlich führt, außer bei „kritischen Infrastrukturen“ (KRITIS), immer Datenschutz. Dabei gilt Datenschutz den Betroffenen, nicht den Organisationen.

- **Datenschutz gründet nicht im individuellen Bedürfnis nach Privatheit.**
  Konzepte wie “Selbstbestimmung”, “Privatautonomie”, „individuelle Grundrechte“ sind funktionale Voraussetzungen für das Funktionieren moderner Gesellschaften. Datenschutz ist ein Indikator für die Moderne einer Gesellschaft.

Objektbereich des Datenschutzes



Datenschutz beobachtet, beurteilt und gestaltet die asymmetrischen Machtbeziehungen zwischen mächtigen **Organisationen** (*Risikogeber*) und im Grundsatz selbstständig agierenden **Personen** (*Risikonehmer*).

Zwischenfrage:
Was sind „*Grundrechte*“?

Grundrechte sind Abwehrrechte für BürgerInnen

(allgemeiner: für Personen) gegen den Staat (allgemeiner: gegen Organisationen).

Datenschutz wacht darüber, dass Grundrechte von Organisationen in der Praxis wirksam umgesetzt werden

(und nicht nur im Grundgesetz oder in der EU-Grundrechtecharta stehen).

***Die* zentrale Regel**
der Datenschutz-Grundverordnung
ist sehr einfach und
kristallinklar formuliert, nämlich:

Organisationen dürfen keine
personenbezogenen Daten
verarbeiten

PUNKT

Und wo steht diese einfache und klare Regel der DSGVO?

„**Verbot mit Erlaubnisvorbehalt**“
Artikel 6 Abs. 1 DSGVO

Organisationen dürfen keine personenbezogenen Daten erheben, verarbeiten und nutzen, es sei denn,

- dass eine **Einwilligung** oder ein **Vertrag** vorliegt, was für den privaten Bereich zentral ist und insbesondere eine klare Darstellung des legitimen Zwecks und der Freiwilligkeit der Erteilung voraussetzt (Einwilligungen können jederzeit zurück genommen werden).
- dass ein **Gesetz** die Verarbeitung regelt, was insbesondere für den öffentlichen Bereich gilt.

Noch eine Bemerkung zum Verhältnis von Datenschutz und IT-Sicherheit („Angreifermodell“)

Die IT-Sicherheit für Computer und Computernetze unterstellt:

- **Jede Person kann ein Angreifer sein!**
  Organisationen überwachen deshalb Personen und zwingen diese nachzuweisen, dass sie keine Angreifer sind.

Was unterstellt der Datenschutz?

- **Jede Organisation ist ein Angreifer!**
  Deshalb müssen Organisationen nachweisen, dass sie keine Angreifer sind, u.a. dadurch dass sie sich nachweislich an Gesetze und Regeln halten und ihre Verarbeitungstätigkeiten und Prozesse sicher beherrschen.

(Hinweis für Kryptologen: Bob (nicht: Eve) ist der wahrscheinlichste und gefährlichste Angreifer für Alice!)

Objektbereich des Datenschutzrechts ist die „Verarbeitung“ einer Organisation

Die DSGVO **regelt nicht**

- das **einzelne personenbezogene Datum,**
- oder **ein technisches Artefakt** (einen Server, eine Applikation oder ein einzelne Komponente) zum Gegenstand einer Prüfung,
- Sondern die **“Verarbeitung“** personenbezogener Daten einer Organisation, die typischerweise auf Informationstechnik aufsetzt (vgl. Art. 4, Abs. 2 DSGVO).
- ***Mit Bezug zur Künstlichen Intelligenz (KI): Aus Datenschutzsicht interessiert nicht ein einzelnes KI-System als technisches Artefakt, sondern ein KI-System ist „nur“ eine (wenn auch besondere) Komponente der Verarbeitung personenbezogener Daten durch Organisationen.***

Agenda

1. Einführung: Was meint “Datenschutz”?
2. Beispiele für intelligente Systeme
3. Wie funktionieren intelligente Systeme?
4. Was ist aus Datenschutzsicht das Problem bei KI/ML?
5. Versprengtes
6. Referenzen

Beispiele für erfolgreiche KI/ML-Anwendungen

Fazit: Maschinen können in komplexen Anwendungsfeldern kognitive Probleme besser lösen als die besten menschlichen ExpertInnen dieser Anwendungsfelder.

IBM „Deep Blue“ schlägt Schach-Weltmeister Kasparov (1996/1997)

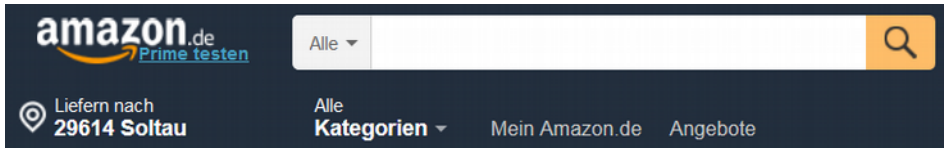
IBM Watson schlägt den Jeopardy-Ratemeister (2011)

Microsoft, Apple, Amazon stellen Sprachassistenz-systeme bereit (2011)

Google: „Alpha Go“ schlägt Go-Weltmeister Lee Sedol (2016)

Carnegie Mellon University: „Libratus“ gewinnt Profi-Pokerturnier (2017)

Beispiele
für erfolgreiche KI/ML-Anwendungen



facebook

Google

Beispiel KI
Facebook heute:
Interview with Entwickler Yann Lecun

Frankfurter Allgemeine
Diginomics

F.A.Z.-INDEX 2.253,53 +0,92 % DAX ® 11.606,33 +1,06 % EUR/USD 1,1488 +0,60 % DOW JONES 25.635,01 --

Q&A WITH YANN LECUN

„Facebook wouldn't work without AI today"

VON ALEXANDER ARMBRUSTER - AKTUALISIERT AM 06.11.2018 - 14:51

Yes. And now, five years later, AI has become very central to Facebooks operations. Facebook today simply wouldn't work without AI or, let's say Deep Learning.

When I log in into my Facebook-Account, where do I meet AI?

First of all, there is all the content that you see and all the content, that you don't see and both are largely determined by AI-Systems. What you see on your newsfeed are pieces of information, that the ML-Algorithms of Facebook have identified as the most likely to be of interest for you. So, there is a model of your taste inside of Facebook...

Beispiele für erfolgreiche KI/ML-Anwendungen

Fähigkeit	Nutzen	FuE-Ansätze	Mögliche Anwendungen	Reife-grad
Gruppen ähnlicher Daten bilden	Diese Fähigkeit ermöglicht bspw. das Erkennen von Mustern und hilft, Strukturen in großen Datenmengen zu erkennen.	▪ Clustering ▪ Tiefe neuronale Netze	▪ Marketing: Kundensegmentierung, Zielgruppenübersicht	1
Objekte klassifizieren	Diese Fähigkeit ermöglicht das Einordnen von Beispielen für die weitere Bearbeitung, um Entscheidungen zu treffen oder Maßnahmen einzuleiten.	▪ Entscheidungsbäume ▪ Stützvektormaschinen ▪ Bayessche Netze ▪ Logistische Regression	▪ Datenfiltersysteme ▪ Sortieraufgaben (z. B. Güteklassifizierung in der Produktion) ▪ Marketing (z. B. Matching von Kunden und Waren)	1
Werte schätzen und vorhersagen	Hier werden lineare oder komplexere Zusammenhänge erkannt und für Vorhersagen über künftige Zustände bzw. Ereignisse genutzt.	▪ Lineare Regression ▪ Regressionsbäume (CART) ▪ Entscheidungsbäume	▪ Generierung von Prognosen (Stau, Angebot- und Nachfrage) ▪ Anomalie-Detektion ▪ Maschinen-/Anlagenoptimierung ▪ Vorausschauende Wartung ▪ Finanz-, Versicherungs- und Rechtswesen ▪ Medizin, Chemie, Materialforschung (Entdeckung neuer Molekularkombinationen etc.) ▪ Steuerungsaufgaben	1
Erfolgversprechende Aktionen für einen Agenten auswählen	Agenten und Roboter können anhand von Feedback lernen, welche Aktionen, Spielzüge etc. die besten Resultate erzielen können. Das ist eine Alternative zum expliziten Planen und Adaptieren von Handlungsfolgen.	▪ Bestärkendes Lernen ▪ Q-Lernen mit tiefen neuronalen Netzen	▪ Robotik (z. B. um das optimale Greifen unterschiedlicher Objekte zu lernen) ▪ Autonomes Fahren (z. B. um in Simulationen erwünschte und unerwünschte Aktionen unterscheiden zu können) ▪ Spieleindustrie ▪ Konsumelektronik	2
Bilder erkennen	Auf Bildern werden Objekte lokalisiert, klassifiziert und ggf. Individuen erkannt.	▪ Tiefe neuronale Netze	Alle Bereiche, in denen Objekterkennung von Nutzen ist, insbesondere ▪ Automotive (Straßenschildidentifikation etc.) ▪ Industrielle Produktion/Industrierobotik ▪ Medien (Suche) ▪ Medizin (Radiologische Diagnostik) ▪ Sicherheit (Videoüberwachung)	1

Legende:

- Stufe 1: Inzwischen gut etabliert
- Stufe 2: Demonstratoren vorhanden, Forschung für komplexere Anwendungen unbedingt erforderlich
- Stufe 3: Noch in früher FuE-Phase

aus: Fraunhofer 2018: S. 33

Definition „Maschinelles Lernen“ (ML)

*„Maschinelles Lernen bezweckt die Generierung von ‚**Wissen‘ aus ‚Erfahrung**‘, indem Lernalgorithmen aus Beispielen ein komplexes Modell entwickeln.*

*Das **Modell**, und damit die automatisch erworbene Wissensrepräsentation, kann anschließend auf neue, potenziell **unbekannte Daten derselben Art angewendet** werden.*

*Immer **wenn Prozesse zu kompliziert sind**, um sie analytisch zu beschreiben, aber genügend viele Beispieldaten – etwa Sensordaten, Bilder oder Texte – verfügbar sind, bietet sich Maschinelles Lernen an.*

*Mit den gelernten Modellen können **Vorhersagen getroffen oder Empfehlungen und Entscheidungen generiert** werden – ganz **ohne im Vorhinein festgelegte Regeln oder Berechnungsvorschriften.**“* (aus: Fraunhofer 2018: 8).

Daraus folgt:
Man kann Computer („KI-Systeme“) konstruieren, bei denen die Konstrukteure *weder den Gegenstandsbereich hinreichend kennen* (und ihn symbolisch deshalb nicht darstellen können) noch die Programmierung des Computers auf der „kognitiven Ebene“ des Computers in Form von *Regeln und Algorithmen beherrschen* müssen.

Das Risiko:
Ein KI-Computer kann beliebige Berechnungen mit personenbezogenen Daten durchführen. Beherrschbarkeit? Zweckbindung? Integrität?

Definition "Künstliche Intelligenz"

>>Wir verstehen „Künstliche Intelligenz“ in diesem Zusammenhang als Sammelbegriff für diejenigen Technologien und ihre Anwendungen, die durch ***digitale Methoden*** *auf der Grundlage potenziell* ***sehr großer und heterogener Datensätze*** *in einem komplexen und* ***die menschliche Intelligenz gleichsam nachahmenden maschinellen Verarbeitungsprozess*** *ein Ergebnis ermitteln, das ggf. automatisiert zur Anwendung gebracht wird. Die wichtigsten Grundlagen für KI als Teilgebiet der Informatik sind die* ***subsymbolische Mustererkennung****, das* ***maschinelle Lernen****, die* ***computergerechte Wissensrepräsentation*** *und die Wissensverarbeitung, welche Methoden der* ***heuristischen Suche, der Inferenz und der Handlungsplanung*** *umfasst.<<*

Datenethikkommission der Bundesregierung 9.10.2018

daten ethik kommission

KI-Typologie
Lernstile, Lernaufgabe, Lernverfahren und Modelle

Lernstil	Lernaufgabe	Lernverfahren	Modell
Überwacht	Regression	Lineare Regression	Regressionsgerade
		Klassifikations- und Regressionsbaum-verfahren (CART)	Regressionsbaum
	Klassifikation	Logistische Regression	Trennlinie
		Iterative Dichotomizer (ID3)	Entscheidungsbaum
		Stützvektormaschine (SVM)	Hyperebene
		Bayessche Inferenz	Bayessche Modelle
Unüberwacht	Clustering	K-Means	Clustermittelpunkte
	Dimensionsreduktion	Kernel Principal Component Analysis (PCA)	Zusammengesetzte Merkmale
Bestärkend	Sequentielles Entscheiden	Q-Lernen	Strategien
Verschiedene	Verschiedene	Rückwärtspropagierung	Künstliche Neuronale Netze

Typische Wahrscheinlich-keitsmodelle von Wissenschaft. (Kausalität)

Zunehmende „Nicht-Trivialität“

Künstliche **N**euronale **N**etze zentrale Technik für *maschinelles Lernen* (ML) (Korrelationen)

aus: Fraunhofer 2018: S. 18

Warum
Techniken der Künstlichen Intelligenz und des maschinellen Lernens einsetzen?

- Komplexe Probleme lassen sich nicht oder nur mit großem Aufwand durch (von Menschen programmierte) Algorithmen lösen.
- Mit starren Algorithmen kann man nur langsam auf sich ändernde (System-) Bedingungen reagieren.
- Wenn Systeme „lernen“ können, können sie aus Fehlern und Erfolgen Schlüsse ziehen, um noch bessere Lösungen zu finden.

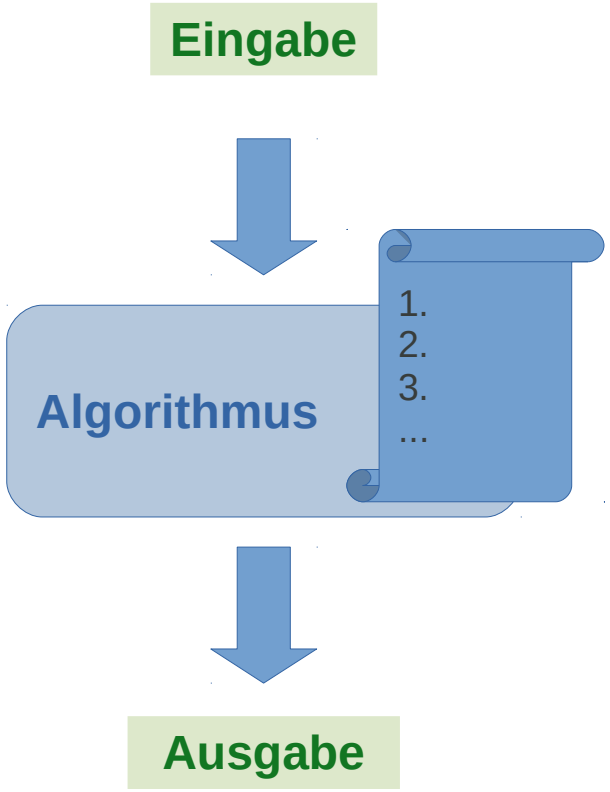
Agenda

1. Einführung: Was meint “Datenschutz”?
2. Beispiele für intelligente Systeme
3. Wie funktionieren intelligente Systeme?
4. Was ist aus Datenschutzsicht das Problem bei KI/ML?
5. Versprengtes
6. Referenzen

Definition Algorithmen

Algorithmen sind eindeutige, endliche Handlungsvorschriften zur Lösung eines Problems

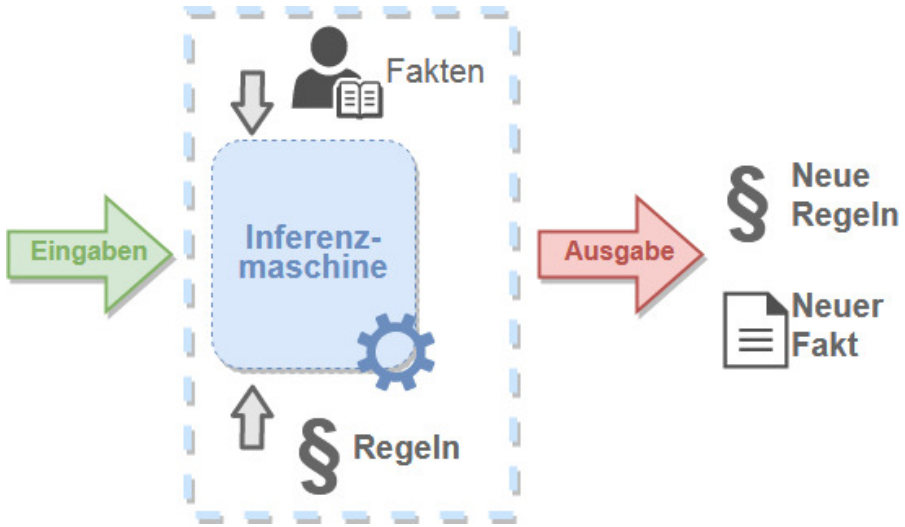
- Von Menschen formulierter / programmierter / regelhafter Lösungsweg
- **Verhalten? Immer gleiche Eingabe führt zu immer gleicher Ausgabe**



<u>Beispiele:</u>

- Theoretische Führerscheinprüfung
- BAföG-Berechnung
- Wahl-Verfahren

Wie funktioniert ein Expertensystem? (Regressionsmodell)



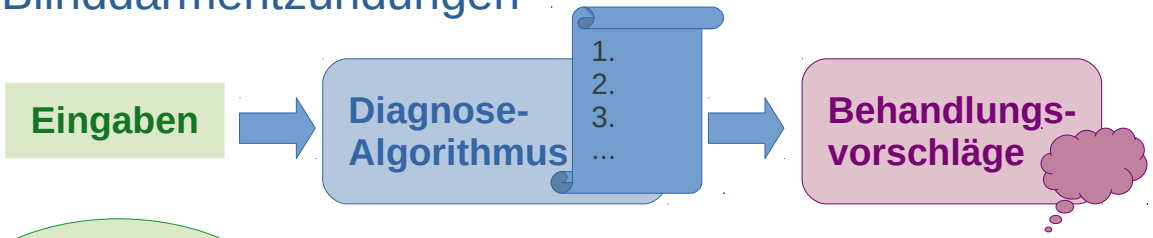
Expertensystem

Neue Regeln und Fakten werden aus bekannten Regeln und Fakten logisch erschlossen

<u>Anwendungsbeispiele:</u>

- Mathematische Problemlösungen
- Medizinische Diagnosen

Beispiel:
LEXMED
Blinddarmentzündungen



Eingabeformular
Abfrage diverser Patientendaten und Symptome

Personenangaben	nicht bekannt	Werte	
Geschlecht	○	● männlich ○ weiblich	?
Altersgruppe	○	○ 0-5 ○ 6-10 ○ 11-15 ○ 16-20 ● 21-25 ○ 26-35 ○ 36-45 ○ 46-55 ○ 56-65 ○ 65-	?
Untersuchungsergebnisse	**nicht untersucht**	**Werte**	
1. Schmerzquadrant	○	○ ja ● nein	?
2. Schmerzquadrant	○	○ ja ● nein	?
3. Schmerzquadrant	○	○ ja ● nein	?
4. Schmerzquadrant	○	○ ja ● nein	?
Abwehrspannung	○	○ lokal ● global ○ keine	?
Loslassschmerz	○	● ja ○ nein	?
Erschütterungsschmerz	○	● ja ○ nein	?
Rektalschmerz	○	○ ja ● nein	?
Darmgeräusche	○	○ schwach ○ normal ● vermehrt ○ keine	?
Sonographisch auffällig	○	● ja ○ nein	?
Pathologisches Urinsediment	○	● ja ○ nein	?
Rektaler Temperaturbereich	○	○ -37.3 ○ 37.4-37.6 ● 37.7-38.0 ○ 38.1-38.4 ○ 38.5-38.9 ○ 39.0-	?
Leukozytenbereich	○	○ 0-6k ● 6k-8k ○ 8k-10k ○ 10k-12k ○ 12k-15k ○ 15k-20k ○ 20k-	?

Wie funktioniert ein Expertensystem? (Regressionsmodell)

```
|
C4.5 [release 8] decision tree generator            Wed Aug 23 13:13:15 :
----------------------------------------
|
    Options:
        File stem <app>
        Trees evaluated on unseen cases
        Sensible test requires 2 branches with >=100 cases
|
Read 9764 cases (15 attributes) from app.data
|
Decision Tree:
|
Leukozyten <= 11030 :
|   Schmerz_bei_Loslassmanoever = 0:
|   |   Temp_re > 381 : 1 (135.9/54.2)
|   |   Temp_re <= 381 :
|   |   |   Lokale_Abwehrspannung = 0: 0 (1453.3/358.9)
|   |   |   Lokale_Abwehrspannung = 1:
|   |   |   |   Geschlecht_(1=m___2=w) = 1: 1 (160.1/74.9)
|   |   |   |   Geschlecht_(1=m___2=w) = 2: 0 (286.3/97.6)
|   Schmerz_bei_Loslassmanoever = 1:
|   |   Leukozyten <= 8600 :
|   |   |   Temp_re > 378 : 1 (176.0/59.4)
|   |   |   Temp_re <= 378 :
|   |   |   |   Geschlecht_(1=m___2=w) = 1:
|   |   |   |   |   Lokale_Abwehrspannung = 0: 0 (110.7/51.7)
```

	Wahrscheinlichkeiten für verschiedene Befunde				
	entzuendet	perforiert	negativ	andere	
	0.568	0.042	0.092	0.298	
App (nicht perf.):OP	0	1000	5800	6000	2366
App.(perf): Not-OP	500	0	6300	6500	2803
NSAP: amb. beob.	12000	150000	0	16500	18017
Sonst.(Op,and.Beh.,weit.Unt.)	3000	5000	2000	0	2097
Stat. beob.	3500	7000	400	600	2496

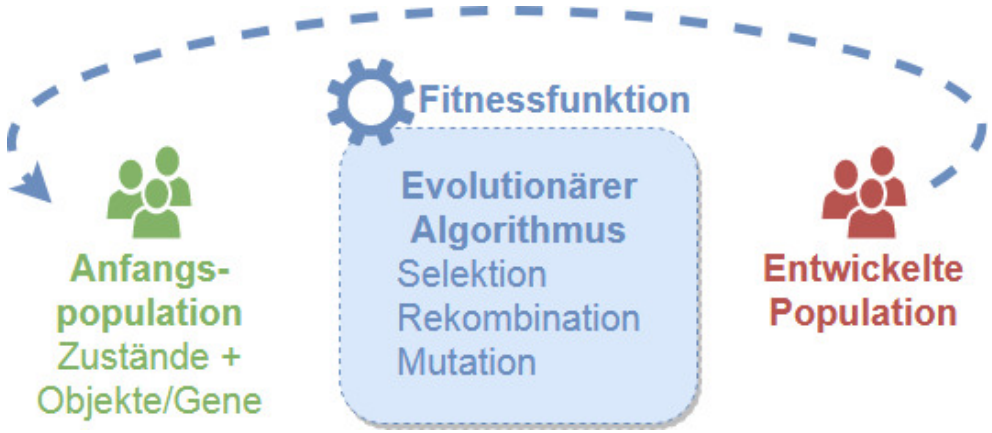
Behandlungsvorschlag: Sonst.(Op,and.Beh.,weit.Unt.)

Mit der Diagnose können Therapie-Optionen bewertet und vorgeschlagen werden.

Auf Grundlage der Eingabedaten und der Verarbeitung anhand eines Entscheidungsbaums wird eine Diagnose erstellt

Wahrscheinlichkeiten für verschiedene Befunde			
entzuendet	perforiert	negativ	andere
0.568	0.042	0.092	0.298

Wie funktionieren evolutionäre Algorithmen?



Evolutionäre Algorithmen

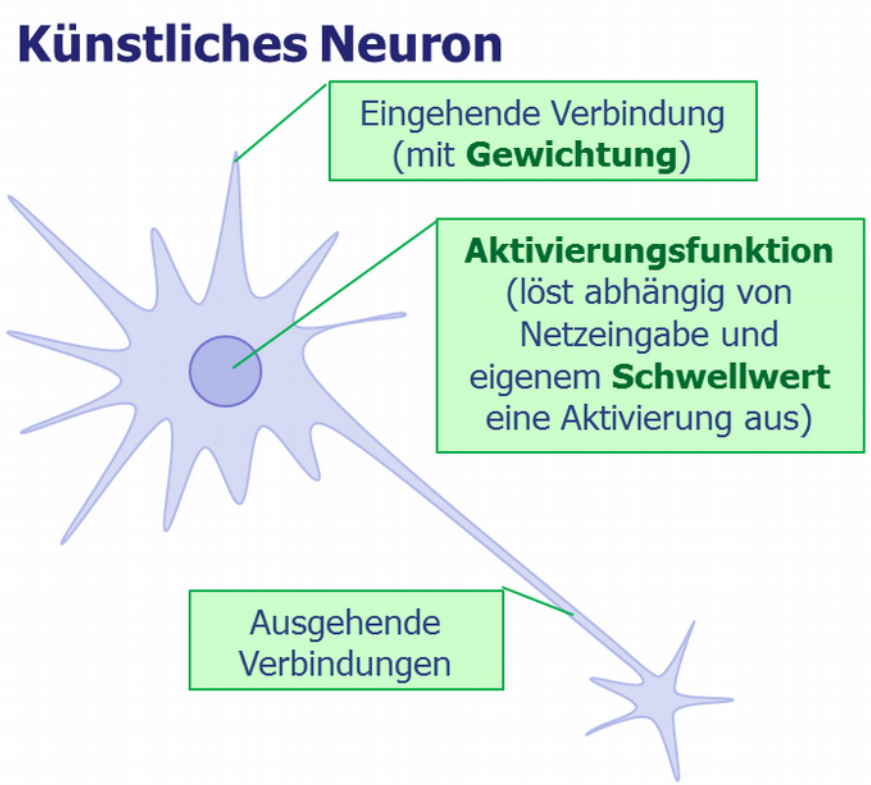
Mithilfe automitisierter Veränderungen, Selektionen und Stabilisierung von Programmcodes können gewünschte Merkmale ausgiebig optimiert werden.

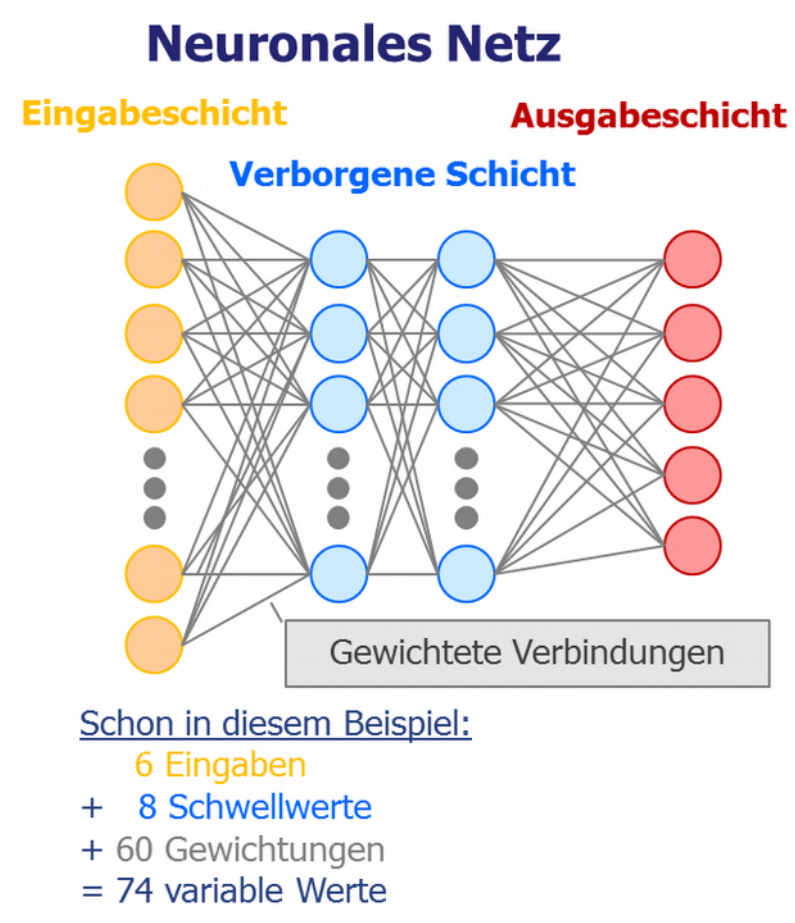
Beispiel:

Die Antenne der Space-Technology-5-Satelliten wurde mit einem Evolutionären Algorithmus entwickelt.

Aus: Wikipedia: https://commons.wikimedia.org/wiki/File:St_5-xband-antenna.jpg
„This file is in the public domain in the United States because it was solely created by NASA. NASA copyright policy states that "NASA material is not protected by copyright unless noted".

Wie funktionieren... Künstliche Neuronale Netze (KNN)?

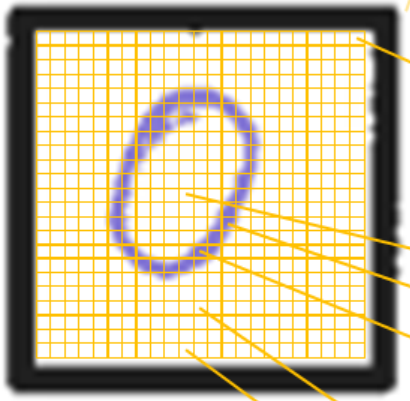




Einige Merkmale künstlicher Intelligenz in neuronalen Netzen:

Lernen erfolgt zunächst auf Trainingsdaten und mit verschiedenen Verfahren (z.B. überwacht, unüberwacht)
Erfahrungen/Wissen ändert sich mit jeder Nutzung. Der Systemzustand ist also ein Berechnungsparameter.
Bewertung einer Lösung erfolgt durch Kostenfunktion, die vorab definiert werden muss.
Nachvollziehbarkeit einer konkreten Berechnung nur schwer möglich, da ein Künstliches Neuronales Netz über eine sehr große Zahl neuronaler Verknüpfungen verfügt.

Wie funktionieren KNN?
Beispiel: Erfolgreiches Erkennen einer geschriebenen Ziffer

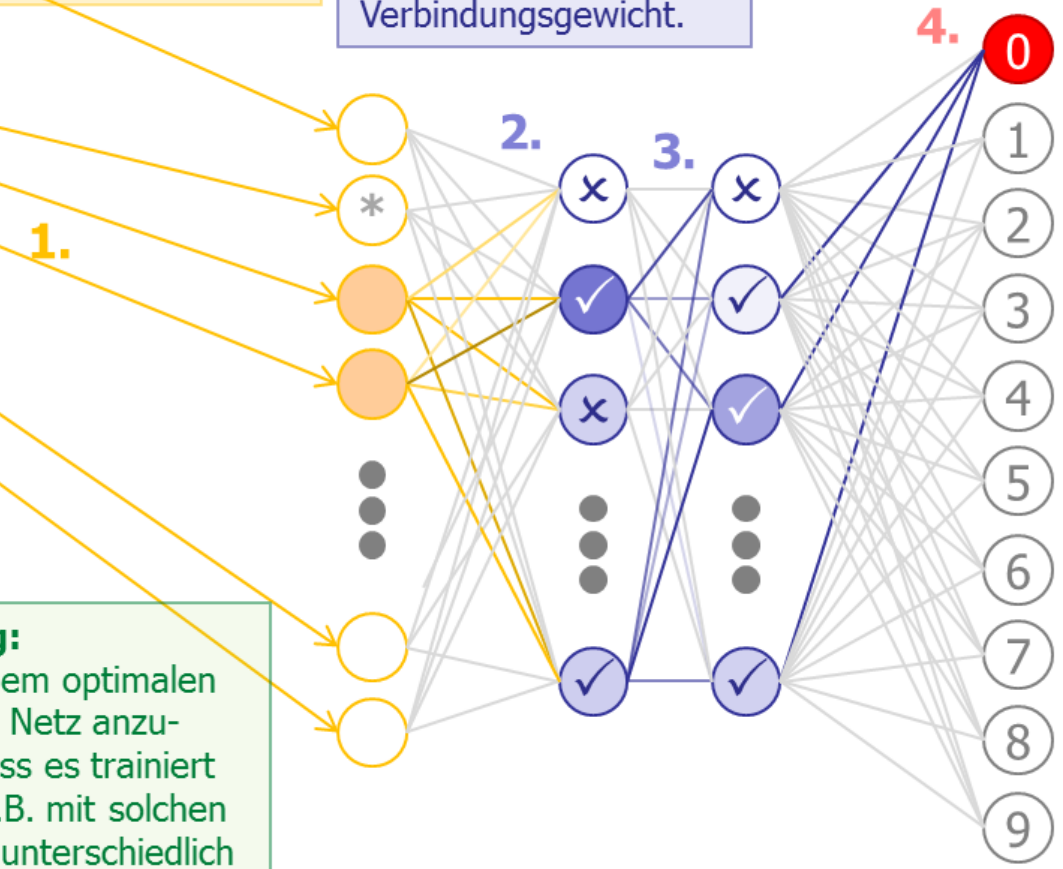


1. Jeder Pixel wird einem Eingabeneuron zugewiesen. Es ist aktiviert, wenn der Pixel farbig ist.

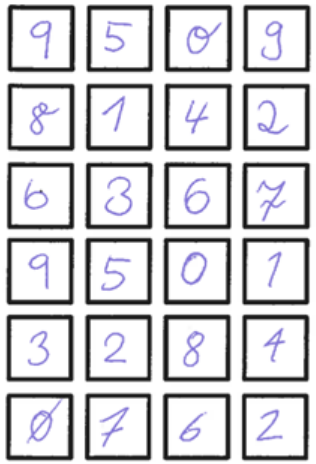
2. Die aktivierten Eingabeneuronen aktivieren weitere Neuronen – abhängig vom Verbindungsgewicht.

3. Sind aktivierte Neuronen (✓) über ihrem Schwellwert aktiviert, aktivieren sie andere Neuronen im Netz.

4. In einem optimalen Neuronalen Netz wird ein Ausgabeneuron aktiviert.



Trainingsdaten:



5. Training:
Um sich einem optimalen Neuronalen Netz anzunähern, muss es trainiert werden – z.B. mit solchen Ziffern, die unterschiedlich geschrieben wurden.

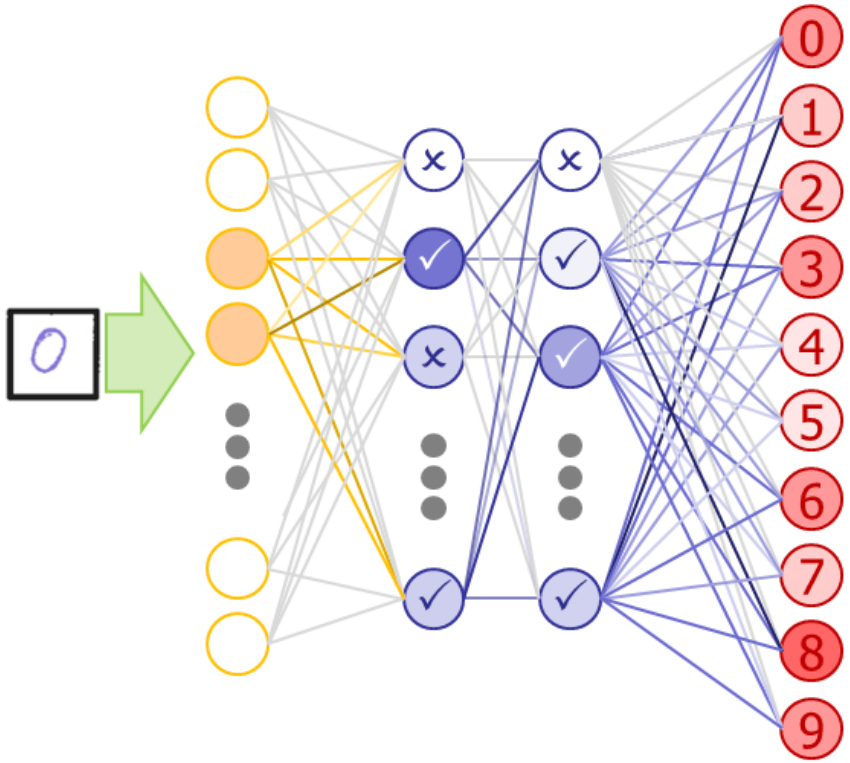
Netzanalyse am Rande:
Zum Beispiel der Pixel in der Mitte und dessen Eingabeneuron (*) sind beim Erkennen einer 0 kaum beteiligt – wohl aber bei anderen Ziffern wie 3, 4, 5, 6, 8 oder 9.

Wie funktionieren KNN?
KNN-Beispiel: Erkennen von geschriebenen Ziffern

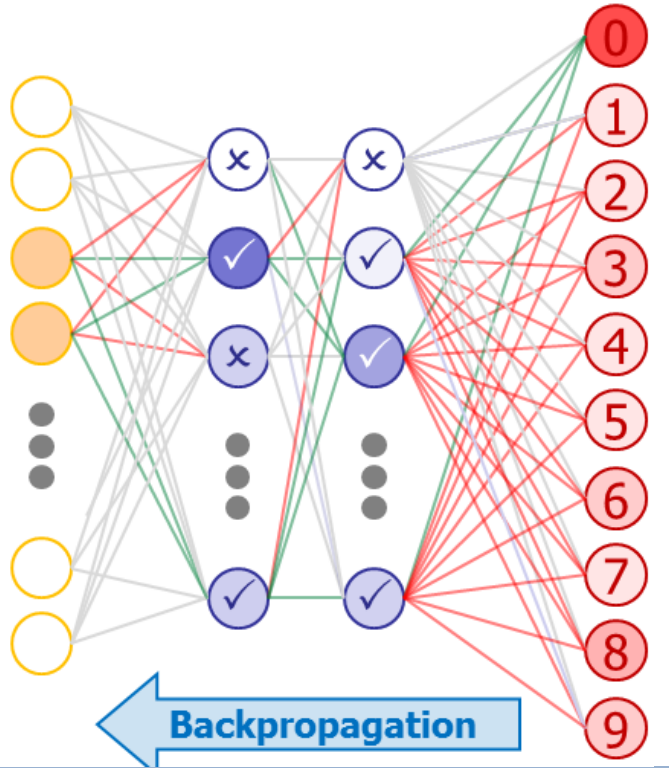
Das Neuronale Netz wird mit einer Trainings-Ziffer aktiviert.

In diesem Beispiel hat das Neuronale Netz eine 8 „erkannt", d.h. das entsprechende Ausgabe-Neuron wurde am stärksten aktiviert.
Nun wird das Neuronale Netz mit der richtigen Lösung trainiert.

Die Gewichte der Verbindungen zwischen Neuronen werden angepasst: Von der 0 ausgehend rückwärts bis zu den Eingabe-Neuronen werden alle aktivierten Verbindungen verstärkt. Entsprechend werden für die anderen Ziffern die Verbindungen geschwächt.
Dieses Lernverfahren wird **Backpropagation** genannt.



0 ist richtig!



Kritik an KI und dem Maschinenlernen (ML)

- **mangelnde Robustheit** gegenüber geringfügigen Transformationen von Trainingsdaten, mit dem besonderen Effekt des „katastrophischen Vergessens“ von schon mal korrekt Abgebildetem durch starke Änderungen von Trainingsdaten;
  (vgl. Fraunhofer 2018: 29)
- **großer Aufwand** für überwachtes Lernen, keine passablen Ergebnisse bei zu wenigen Trainingsdaten;
- es **fehlen garantierte Vertrauensniveaus**, die es gestatten würden, dass Verantwortliche berechenbare Vertrauens- oder Angreifermodelle bzgl. des KI-Systems formulieren.
- **Mangel an verständlichen Erklärungen und Begründungen** zu Lernergebnissen gegenüber den Nutzern bzw. Betroffenen;

- Fazit Laßmann: “***Deep-Learning-Systeme sind also nicht prüfbar, nicht evaluierbar, ändern Ihre Eigenschaften, liefern keinerlei Begründung, sind leicht zu manipulieren, willkürlich aufgebaut und immer ungenau.***“ (vgl. Laßmann 2018, Pos. 1304)

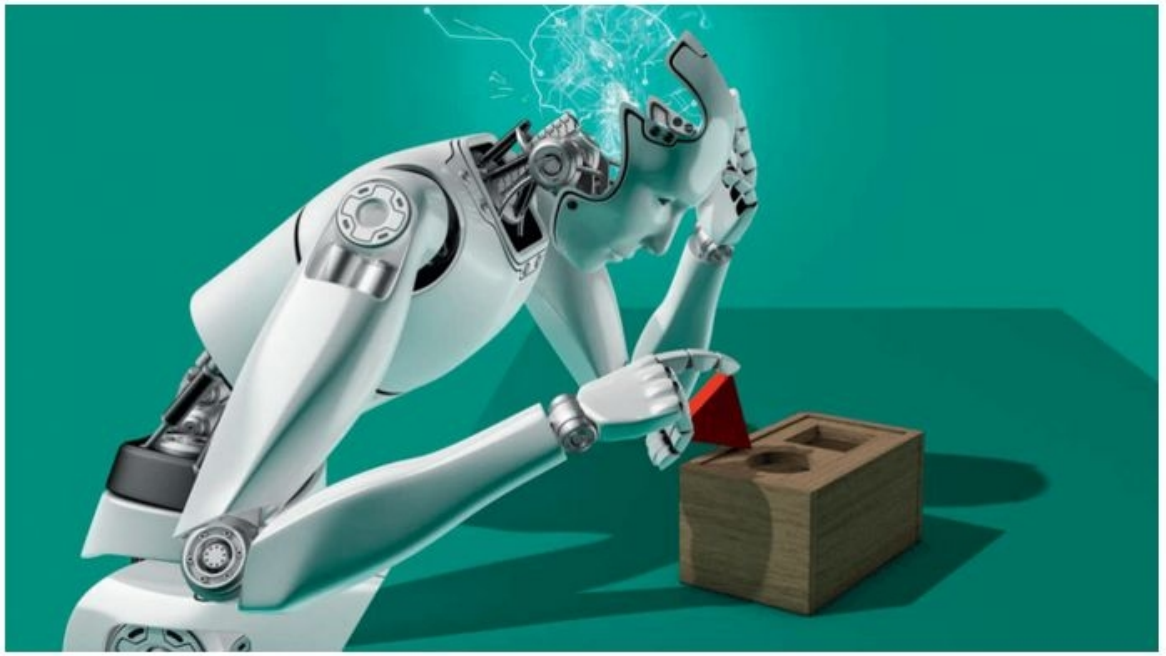
Kritik an KI
Heise-Meldung vom 9.11.2018

08:00 Uhr | c't Magazin

Künstliche Intelligenz: Dümmer als man denkt

Künstliche neuronale Netze sprechen mit Menschen und gewinnen Strategiespiele. Mit Denken und Intelligenz hat das noch wenig zu tun, erklärt die aktuelle c't.

Von **Andrea Trinkwalder** 4



„(…) eine Fehlerrate von nur einem Prozent kann für die eine Anwendung hervorragend, für die andere inakzeptabel sein. (…) Außerdem erkennen künstliche neuronale Netze nur Muster, keine kausalen Zusammenhänge. (…) Die meisten Experten sind sich einig, dass ein fundamentaler Durchbruch in der KI nötig ist, um menschliches Denken auch nur ansatzweise zu simulieren.“

Quelle: https://www.heise.de/newsticker/meldung/Kuenstliche-Intelligenz-Duemmer-als-man-denkt-4216533.html

Agenda

1. Einführung: Was meint “Datenschutz”?
2. Beispiele für intelligente Systeme
3. Wie funktionieren intelligente Systeme?
4. Was ist aus Datenschutzsicht das Problem bei KI/ML?
5. Versprengtes
6. Referenzen

Wie prüft man Datenschutz?

- Funktionieren KI-Systeme wie KNN **hinreichend verlässlich**? Sind sie “austrainiert”? Wann gilt ein KI-System als austrainiert?
- Wie steht es um die **Verantwortung von oder für KI-Systeme,** insbesondere wenn sie falsch entscheiden?
- Und dann: **Wie prüft man KI-Systeme, ob sie funktional und rechtlich korrekt funktionieren**?

Verantwortlichkeit “Elektronische Person”?


Quelle: https://www.rubikon.news/artikel/kunstliche-intelligenz-als-gefahr

WIKIPEDIA
Die freie Enzyklopädie

Hauptseite
Themenportale
Zufälliger Artikel

Mitmachen

Artikel verbessern
Neuen Artikel anlegen

Wikipedia durchsuchen

Elektronische Person


https://www.datacenter-insider.de/kuenstliche-intelligenz-haelt-im-it-service-management-einzug-a-745300/

KI und Verantwortlichkeit

- Option 1: **Ein KI-System ist für sich selber verantwortlich.**
  - Ein KI-System ist kein Zweck an sich selbst im Sinne eines Subjekts.
- Option 2: Der **Käufer/Nutzer** eines KI-Systems ist für die Aktivitäten seines KI-Systems verantwortlich.
  - Dann muss der Käufer/Nutzer auch die Vollkontrolle über das System innehaben. Wenn er die nicht hat, ist diese Option hinfällig.
- Option 3: Der **Betreiber einer KI** ist für ein Verfahren, in dem eine KI zum Einsatz kommt, und damit auch für die KI selber, verantwortlich.
  - Das ist der klassische datenschutzrechtliche Fall eines Verantwortlichen, der eine Technik, oder auch intelligente MitarbeiterInnen, innerhalb seiner Verarbeitung nutzt.
- Option 4: Der **Hersteller / Customizer (Trainer)** eines KI-Systems ist verantwortlich.
  - Hersteller und Customizer haben die größte Kontrolle über Ziele und Mittel des KI-System, deshalb können diese noch am ehesten die Verantwortung übernehmen.
- Vermutliche Lösung: **Geteilte Verantwortung** durch Hersteller / Customizer / Betreiber / Nutzer entsprechend unterschiedlicher Risiko- oder Fehlertypen.
  (Ein Eigentümer kann eine KI wahrscheinlich zumindest in den Pionierzeiten auch zu kriminellen Aktivitäten einsetzen, dann Direktverbindung des KI-Systems zur Polizei?)

Anforderungen des Datenschutzes, Datenschutz-Risiken bzgl. KI: Artikel 5 DSGVO

Art. 5 Abs. 1 DSGVO „Personenbezogene Daten müssen“
(a) „... in einer für die Person nachvollziehbaren Weise verarbeitet werden ... (**Transparenz**).“
(b) „... für festgelegte eindeutige und legitime Zwecke erhoben werden ... (**Zweckbindung**).“
(c) „... auf das für die Zwecke der Verarbeitung notwendige Maß beschränkt sein (**Datenminimierung**).“
(d) „... damit personenbezogene Daten, die im Hinblick auf die Zwecke der Verarbeitung unrichtig sind, ... **unverzüglich gelöscht oder berichtigt** werden.“
(f) „... **Schutz vor Verlust** ... **Integrität und Vertraulichkeit**“.

Schutzziele

- Transparenz
- Nichtverkettung
- Intervenierbarkeit
- Verfügbarkeit
- Integrität
- Vertraulichkeit

Die Risiken einer KI aus Sicht des Datenschutzes?
Wenn Organisationen KI nutzen, und **gegen diese 6 Schutzziele bei ihrer Datenverarbeitung verstoßen**.

Referenzmaßnahmen zur Umsetzung der Schutzziele (Standard-Datenschutzmodell (SDM))

SDM
Standard-Datenschutzmodell

V 1.1- 2018

Sicherstellung von ***Verfügbarkeit***
Redundante Datensätze, IT-Systeme, Prozesse, „schnelle Reparaturzeiten“

Sicherstellung von ***Integrität***
Hash-Wert-Vergleiche, Härten von IT-Systemen, Festlegen von Min./Max.-Referenzen bei Prozessen, Steuerung der Regulation von Prozessen

Sicherstellung von ***Vertraulichkeit***
Verschlüsselung, Rollen- und Berechtigungskonzepte

Sicherstellung von ***Transparenz***
Prüffähigkeit durch Spezifikation, Protokollierung, Dokumentation, Tests und Freigaben

Sicherstellung von ***Nichtverkettbarkeit*** durch Zweckbestimmung/-bindung, Pseudonymität, Anonymität; Trennung und Isolierung von Datenbeständen, IT-Systemen, Prozessabläufen, Rollen- und Berechtigungskonzepte

Sicherstellung von ***Intervenierbarkeit***
SPOC für Änderungen, Korrekturen, Löschen, Aus-Schalter, standardisierte Changemanagementprozesse in Organisationen

Typische Angreifer aus Sicht des Datenschutzes

- Die Organisation, die personenbezogene Daten verarbeitet
- KI-Kontext: Hersteller, Customizer, Trainer, Datenmodellierer
- HW/OS-Hersteller
- Sicherheitsbehörden
- Leistungsverwaltung
- Bereitsteller von IT-(Infrastruktur)Diensten
- Bereitsteller kritischer Infrastrukturen (wie bspw. Energieversorger, Mobilfunk)
- Versicherungen und Banken
- Forschungsinstitute, insbes. psychologischer und sozialwissenschaftlicher Art
- Krankenhäuser, Ärzte, Rechtsanwälte
- Startups, Werbeagenturen
- Untätige Aufsichtsbehörden
- Cracker

Datenschutz-Risiken
3 Typen

- Risikotyp 1 „Grundrechtseingriff“ zu intensiv:
  Die Datenverarbeitung einer Organisation ist unfair, hält sich nicht an das Gesetz, wird nicht durch einen legitimen Zweck begrenzt, die Anforderungen der DSGVO werden nicht wirksam umgesetzt.
- Risikotyp 2 „Schutzmaßnahmen des Datenschutzes“ versagen,
  bspw. wenn Daten ohne Zweckbindung beliebig verarbeitet werden oder die Protokollierung der Aktivitäten lückenhaft ist oder kein Datenschutz-Controlling implementiert ist.
- Risikotyp 3 „Schutzmaßnahmen der Informationssicherheit“ versagen,
  bspw. wenn Zugriff durch Unbefugte auf personenbezogene Daten besteht oder deren Korrektheit infrage steht.

Kurzes Zwischenfazit

1. KI-Systeme sind Teil einer Verarbeitung personenbezogener Daten einer Organisation. **Es gibt keine autonom in der Welt daseiende, selbstverantwortliche KI-Systeme analog zu Menschen.**
2. Für den Betrieb von KI-Systemen sind vor allem die **Hersteller und Customizer sowie die Betreiber verantwortlich**, und für bestimmte Typen für Anwendungen eines KI-Systems, die Eigentümer bzw. Nutzer (Stichwort: KI-Assistenz für strafbare Handlungen).
3. Diese Verarbeitungen mit KI müssen **den Grundrechten bzw. den Anforderungen der DSGVO genügen**. Artikel 5 DSGVO nennt die Anforderungen, die übersetzt in 6 Schutzziele sich mit Schutzmaßnahmen umsetzen lassen.
4. *Inwieweit helfen nun die Maßnahmen des Datenschutzes (z.B. aus dem Katalog des SDM), um die Risiken künstlicher Intelligenz und maschinellen Lernen zu verringern und ein angemessenen Schutzniveau zu erreichen?*

Funktionale Anforderungen an KI- und ML-Systeme

Die Verfahren, in denen Organisationen KI-Systeme zur Verarbeitung personenbezogener Daten nutzen, müssen den Grundsätzen aus Artikel 5 DSGVO genügen. Das heisst, dass diese Systeme

- **zweckgebunden**
- **intervenierbar**
- **verfügbar**
- **integer**
- **vertraulich**
- **transparent**

funktionieren müssen. Wenn das garantiert und überprüfbar wirksam ist, dürfen KI-Systeme eingesetzt werden.

Schutzmaßnahmen auf der konventionellen („nicht-kognitiven”) Ebene der IT

KI-Systeme müssen konventionell sicher auf der Ebene der IT-Funktionen betrieben werden. Das bedeutet z.B.

- **redundante Systeme**, Ausfallzeiten, Reparaturvereinbarungen;
- **Härten** der Systeme (keine undefinierten Dienste, Nutzer, Zugriffsrechte);
- Zertifikate bzgl. **Verschlüsselung und Integritätssicherung**, Authentizität aller Beteiligten und Systeme, der Interaktions- und Kommunikationskanäle;
- Umsetzen der Nachweispflichten durch Dokumentation und **Protokollierung** der Aktivitäten auf sämtlichen Layern;
- Interventionsmöglichkeiten für Betroffene; **Aus-Schalter**;
- **Pseudonymisierung und Anonymisierung** von Daten, sofern das KI-System an externe Systeme mit einer anderen Zweckbindung angeschlossen ist.

Alle Schutzmaßnahmen müssen dabei **zumindest hohen Schutzbedarf** erfüllen.

- Legitime Zwecksetzung und Zweckdefinition: **Welchen Zweck soll das KI-System erfüllen?** Zweckabgrenzung, Gebot der vertikalen/horizontalen Zweckbindung; Definition was als Diskriminierung ausgeschlossen werden soll;
- Wer **darf den Zweck eines KI-Systems bestimmen**?
- **Die Datenkuration für maschinelles Lernen muss zweckbestimmt sein:**
  - nur die theoriegestützt **domänenspezifisch relevanten Daten** sind zum Training heranzuziehen, entsprechend sind zutreffende Datenquellen aus der Wissensdomäne auszuwählen;
  - **Ein Datenbestand ist zu komplettieren** und **fehlerhafte Daten sind zu entfernen**;
  - **Dimensionalität** der Daten ist zu reduzieren, Ausprägungen sind zu skalieren.

Eine KI-Entscheidung ist gültig, wenn sie sich zugleich im Rahmen des Erwartbaren der Nutzer und des normativ Geforderten befindet. Beide Erwartungslagen wurden idealerweise im Vorhinein explizit gemacht.

- Das **dem Zweck angemessene Lernmodell** ("Stand der Technik") muss bestimmt, programmiert oder trainiert werden.
- Durch permanente Tests (auch Provokationen!) sicherstellen, dass spezifizierte, **antrainierte Eigenschaften nicht verloren** gehen.
- Generell sollte von riskanten KNN-Modellen zu Regressionsmodellen zurückgekehrt werden (Wechsel von Modellen der **Korrelation zu Kausalitätsmodellen)**;
- Einrichtung von **Sicherheitskorridoren**, in denen Störungen und Probleme der KI zu keinen überraschenden Verhaltensänderungen gemäß Zwecksetzung führen (Resilienz-Aspekt).
- **Prüf- und Genehmigungspflicht** für KI-Systeme mit Personenbezug installieren.

Schutzziel Vertraulichkeit

- Die Trainings und die Verarbeitung der Trainigsdaten sollten in einem separiert-geschütztem **Container für IT-Prozesse und Daten** durchgeführt werden.
- Roh-Daten aus der Sensorik eines laufenden Betriebs einer KI sollten weder ungesichert gespeichert noch unbearbeitet bspw. an eine generalisierte KI-Modellierung, etwa beim Hersteller des KI-Systems oder bei einem KI-Betreiber einer bestimmten Wissens-Domäne, übermittelt werden. ("**Google- und Schufa-Problematik**")
- Weitere Angreifer von KI-Daten, wer darf befugt zugreifen auf insbesondere privat erzeugte KI-Daten: Auch die latent hoch-interessierten Sicherheitsbehörden, Banken und Versicherungen, Forschungsinstitute? ("**Hartz4- bzw. Abhängigen-Problematik**")

Schutzziel Intervenierbarkeit

- Bei KI stellt sich die Gretchenfrage einer jeden Vollautomation: **Toppt in einer Konfliktsituation der Pilot die Maschine (KI) oder die Maschine den Piloten?** Die generelle Antwort des europäischen Datenschutzes lautet: Der Verantwortliche und/oder vielfach gerade der betroffene Mensch muss bei einer Maschine, zumindest sofern diese in ihren Folgen allein ihm assistiert oder ihn betrifft, eingreifen können. Frage dann, auf welcher Ebene eines Systems und wie unmittelbar ein Eingriff möglich sein muss.
- Bei automatisierten Entscheidungen: Entweder viele versichernde **Rückfragen an den Anwender** der KI (analog Einwilligung) oder an eine zweite unabhängige Kontroll-KI(?) stellen.
- **Interventions-Skala** (Datenverarbeitung Lenken, Ändern, Stoppen) für laufenden Betrieb: Von “Beschwerdeformular” über “Freigabe durch Experten”, “Ausknopf” bis “Totmann”.
- Nutzer kontrolliert selber, welche Daten er zum Training der KI freigibt und welche nicht (“**nutzerkontrolliertes Kuratieren**”).
- Jederzeitigen **Verzicht** auf KI-Assistenz sicherstellen, ohne dass deshalb für das System ein “Notfall” besteht oder der nutzenden Person durch den Betreiber besondere Haftungsrisiken aufgebürdet werden.

Schutzziel Verfügbarkeit

Beim Ausfall des KI-Systems:

- Übernahme der KI-Funktionen durch externes **Ersatz-KI-System** (autonomes KfZ ohne Führerschein des Nutzers) oder durch **Nutzer** ermöglichen.
- **Jederzeitigen vollständigen Verzicht** (bspw. auf Sprach-Assistent) sichern.
- **Vollständige Selbststeuerung** (bspw. bei einem autonomen KfZ sofern Nutzer Führersschein hat, “Handbetrieb”) ermöglichen.
- Teilausfall: **Notbetrieb** oder **Hybridsteuerung** (Nutzer agiert als Assistent des KI-Systems) oder **Fernsteuerung**
- *Aus Datenschutzsicht: Muss auch bei komplexen Systemen immer die Option auf “Selbststeuerung” angeboten werden?*

Schutzziel Transparenz 1/2

Nicht Haftungsfragen sondern **die Prüfbarkeit, ob die Intensität des Grundrechtseingriffs** in die Autonomie betroffener Personen durch eine KI auf dem geringst möglichen Eingriffsniveau und dem größtmöglichen Nutzen für den Nutzer ausgerichtet ist, steht im Vordergrund der datenschutzrechtlichen Transparenzanforderungen.

Zweck: Herstellen der **Prüfbarkeit** (= Soll-Ist-Bilanz) bzgl. der vorgenannten materiellen Eigenschaften des KI-Systems insbesondere anhand der Protokolle (Selbstauskunftsfähigkeit?) und Nachvollziehbarkeit von **KI-Entscheidungen** und **Berechnungen**.

Schutzziel Transparenz 2/2

Konkret: KI-Systeme bedürfen der **Spezifikation, Protokollierung und Dokumentation:**

- welche **KI-Komponenten** zum Einsatz kommen,
- der **menschlichen Beteiligung** an den Entscheidungsfindungen innerhalb einer Verarbeitung;
- der **Herkunft der Daten**;
- der **Form des Kuratierens** (Definiern, Sammeln, Selektieren, Umwandeln, Verifizieren) und Anreicherung der Rohdaten zu Modell- oder Trainingsdaten;
- des **Lernstils** (überwacht, unüberwacht, bestärkend);
- des verwendeten **Lernmodells** (von Regressionsmodell bis KNN);
- der **Organisationen**, die die Komponenten des KI-Systems hergestellt und über die Auswahl, Konfiguration, Implementation und Betrieb der verwendeten KI-Technik, das Kuratieren der Daten, das Training und der Auswahl der Modelle entschieden haben.
- ein **Gutachten zur Vollständigkeit der Repräsentativität der Wissensdomäne** die die Ki beherrschen soll (und die sich historisch ändert);
- die **Implementation des KI-Algorithmus**, insbesondere der regelbasierten Instruktionen und Entscheidungen;
- Inwieweit die von der DSGVO geforderten **Maßnahmen des Datenschutzes und der IT-Sicherheit** sowie **Regeln der Nichtdiskriminierung** umgesetzt wurden.

Agenda

1. Einführung: Was meint “Datenschutz”?
2. Beispiele für intelligente Systeme
3. Wie funktionieren intelligente Systeme?
4. Was ist aus Datenschutzsicht das Problem bei KI/ML?
5. Versprengtes
6. Referenzen

Anstelle eines Fazit
Versprengte Ideen

- Wahrscheinlich werden **Hybrid-Systeme** zum Einsatz kommen, in denen Menschen und KI-Systeme zu beiderseitigem Nutzen zusammenarbeiten.
- **KNN-KI-Systeme müssen vermutlich (auch) durch KNN-KI-Systeme kontrolliert werden.** Kein Betrieb einer KI ohne eine explizite Datenschutz-Policy mit einer datenschutzgetriebenen Prüf-KI als Bestandteil eines Datenschutz-Managementssystems. **Dies setzt Unabhängigkeit der Kontroll-KI-Systeme von den zu kontrollierenden KI-Systemen voraus** und die Hoheit der Customizer (sprich: Aufsichtsbehörden) über die Kuratierungs- und Trainingsprozesse. (IBM: “trusted ai-services”, https://newsroom.ibm.com/IBM-watson?item=30657)
- Das **politisch** zu lösende Problem besteht darin dafür zu sorgen, dass es **viele unterschiedliche, also unabhängige, KI-Hersteller und System-Customizer** gibt.
- Rechtlich ist **Beweislastumkehr** zu fordern: Die Beweislast bei Risiken muss in der Regel bei den Herstellern, Customizern und Betreibern eines KI-Systems liegen.
- Verarbeitungen mit KI-Komponenten vom Typ “Künstliche Neuronale Netze” und “deep learning” müssen von Aufsichtsbehörden vor deren Einsatz **geprüft und freigegeben** werden.
- **Sprachpolitik und Framing**: Der Vergleich der KI zur menschlichen Intelligenz leitet vollständig fehl, es wird bei KI kein „Hirn“, schon mal gar kein menschliches, nachgebaut. Es handelt sich bei KI und ML um flexible Anpassungsautomaten im Kontext der Vollendung der Automation.
- These: Das nunmehr weltweit vernetzte “maschinelle Lernen” und “künstliche Intelligenz” bilden den **Abschluss der industriellen Revolution**. Datenschutz weist dabei die konkreten Maßnahmen und Strategien aus, damit bei diesem Prozess der Durchtechnisierung der Organisationen die Autonomie und Individualität von Menschen durch Organisationen in der modernenen Gesellschaft nicht gebrochen wird.

Ach ja…
und was sagt eigentlich die DSGVO zur KI?

DSGVO / Artikel 22 - "Automatisierte Entscheidungen im Einzelfall einschließlich Profiling"

(1) **Die betroffene Person hat das Recht, nicht einer ausschließlich auf einer automatisierten Verarbeitung** — einschließlich Profiling — beruhenden Entscheidung **unterworfen zu werden**, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.

(2) **Absatz 1 gilt nicht, wenn die Entscheidung**

a) für den Abschluss oder die Erfüllung eines Vertrags zwischen der betroffenen Person und dem Verantwortlichen erforderlich ist,

b) (…) oder

c) **mit ausdrücklicher Einwilligung der betroffenen Person erfolgt**.

(3) (...)

Agenda

1. Einführung: Was meint “Datenschutz”?
2. Beispiele für intelligente Systeme
3. Wie funktionieren intelligente Systeme?
4. Was ist aus Datenschutzsicht das Problem bei KI/ML?
5. Versprengtes
6. Referenzen

Literatur

- Bundesregierung 2018/07: „**Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz**“ https://www.bmbf.de/files/180718%20Eckpunkte_KI-Strategie%20final%20Layout.pdf
- Datenethikkommission: „Empfehlungen der **Datenethikkommission** für die Strategie Künstliche Intelligenz der Bundesregierung“; https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2018/empfehlungen-datenethikkommission.html
- Diakopoulos, Nicholas / Deussen, Oliver, 2017: Brauchen wir eine **Rechenschaftspflicht** für algorithmische Entscheidungen? in: Informatik-Spektrum, Ausgabe 4: 362-366
- DSBK 2018: **Standard-Datenschutzmodell**, Handbuch, V1.1; https://www.datenschutzkonferenz-online.de/
- Fraunhofer 2018: **Maschinelles Lernen** – Eine Analyse zu Kompetenzen, Forschung und Anwendung; https://www.bigdata.fraunhofer.de/de/big-data/kuenstliche-intelligenz-und-maschinelles-lernen/ml-studie.html
- Laßmann, Günter, 2018: Asimovs **Robotergesetze** – Was leisten sie wirklich? Telepolis.
- Pohle, Jörg, 2018: **Geschichte und Theorie des Datenschutze**s; https://edoc.hu-berlin.de/handle/18452/19886
- Ramge, Thomas, 2018: **Mensch und Maschine**: Wie Künstliche Intelligenz und Roboter unser Leben verändern, Reclam.
- Rost, Martin, 2018: **Künstliche Intelligenz**; in: DuD- Datenschutz und Datensicherheit, 37. Jahrgang, Heft 9: 558-565.

Vielen Dank für Ihre Aufmerksamkeit!

ULD
Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein

Martin Rost / Benjamin Walczak
Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein
Telefon: 0431 988 – 1200
uld32@datenschutzzentrum.de
http://www.datenschutzzentrum.de/