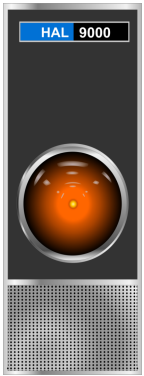
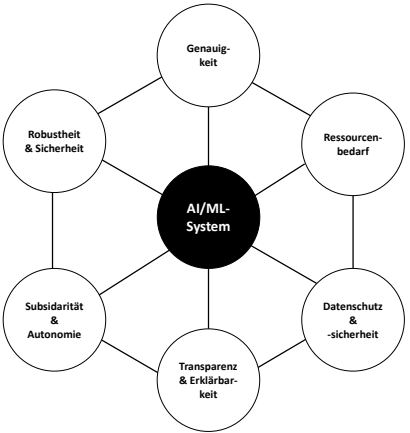
Anforderungen an ML-Systeme im praktischen Einsatz



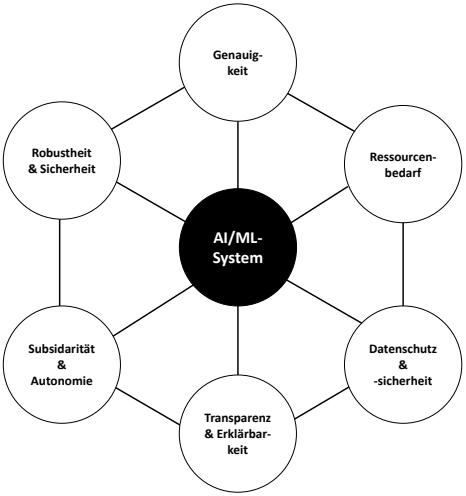
Prof. Dr. Jörg Frochte

Maschinelles Lernen



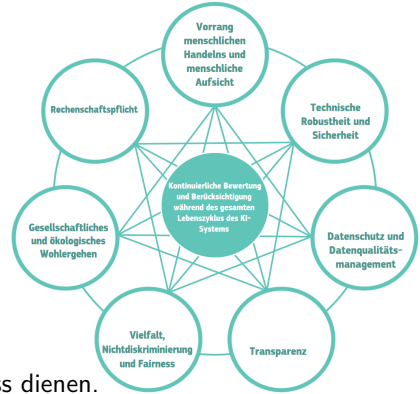
Mehrdimensionale Anforderungen (Frochte, 2018/19/20)

- Genauigkeit ist wichtig, aber abseits von Kaggle Competitions nicht das einzige Maß.
- Oft ist ML ein wichtiger Baustein eines Produktes, aber nicht das ganze Produkt.
- Die Abb. zeigt ein Spannungsfeld von Anforderungen.
- Ob Begriffe gegenüber liegen oder nicht, hat keine Semantik.
- Das Verhältnis zwischen den Anforderungen ist komplex und teilweise verfahrensabhängig.
- Was wie wichtig ist, hängt vom Einsatzgebiet ab.



Vertrauenswürdige KI bzw. Trustworthy AI (EU Kommission, 2019)

1. Menschliches Handeln, die Grundrechte und die letztlich menschliche Beaufsichtigung müssen gewährleistet sein.
2. KI muss technische **Robustheit und Sicherheit** bieten.
3. **Datenschutz und Datenverwaltung** gelten auch und erst recht für KI.
4. Die Funktion von KI-Systemen und insbesondere ihre jeweilige Entscheidungsfindung sollen **transparent** und rückverfolgbar sein.



5. KI-Systeme sollen Vielfalt, Nichtdiskriminierung und Fairness dienen.
6. Das ökologische & gesellschaftliche Wohlergehen muss bei KI-Systemen höchste Priorität haben.
7. Ergebnisse des KI-Einsatzes müssen **überprüfbar** sein, negative Auswirkungen sollen minimiert, Rechtsmittel und Verantwortbarkeit sichergestellt sein.

Transparenz und Erklärbarkeit

- Das Problem ist, dass die Entscheidungen, die ein ML-Verfahren trifft, oft auf ungewöhnlichen Grundlagen fallen.
- Ein bekanntes Beispiel ist die Fehlklassifikation dieses Huskys als Wolf. Grund war der Schnee auf dem Bild.

(a) Husky classified as wolf

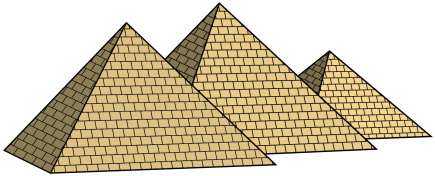
(b) Explanation

- Hier im Einsatz: Eine spezielle Klasse von neuronalen Netzen
- Die Netze liegen sehr oft richtig, aber wenn es einen Fehler gibt, ist dieser oft dramatischer als bei einem Menschen.
- Ein Mensch verwechselt vielleicht eine Biene und eine Wespe, aber er kommt nicht auf die Idee, eine ganz andere Tierart zu benennen.
- Auch ist sich der Mensch (oft) ggf. vorhandener Unsicherheiten bewusst.

Quelle: *Why Should I Trust You?: Explaining the Predictions of Any Classifier* von M. T. Ribeiro, S. Singh & C. Guestrin (2016) in Proc. of the 22nd ACM SIGKDD

Ressourcen

- Training und Verwendung kosten
  Rechenleistung = Energie (Minimierung)
- Training und Verwendung kosten
  Zeit (Minimierung)
- Bei Eager Learnern ist das Training aufwendig, bei Lazy Learnern verlagert es sich zur Ausführung.
- Einige Verfahren sind komplexer zu parametrieren und aufwendiger zu entwickeln und zu warten.



You can do anything you set your mind to when you have vision, determination, and an endless supply of expendable labor.

Anonymisierung und Pseudonymisierung

Anonymisierung (§ 3 Abs. 6 BDSG)

das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können.

Pseudonymisierung (§ 3 Abs. 6a BDSG)

das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren.

IP-Adressen sind keine Pseudonymisierung oder Anonymisierung

Laut EuGH (2016) ist die IP-Adresse ein Merkmal, welches zu personenbezogenen Daten führt; jedenfalls für alle Länder, in denen die rechtliche Möglichkeit besteht, den Inhaber anhand weiterer Zusatzinformationen, über welcher der Provider des Nutzers verfügt, zu bestimmen.

Datenschutz und Datensicherheit

- Der Name einer Person interessiert uns beim Training maschineller Lernalgorithmen so gut wie nie.
- Eine einfache Pseudonymisierung kann recht leicht mittels z.B. einer UUID vorgenommen werden.
- Will man z.B. in einer Cloud Daten verarbeiten, so kann die Zuordnung lokal (verschlüsselt) gespeichert und die Verarbeitung auf den pseudonymisierten Daten durchgeführt werden.
- Je reicher an Merkmalen unsere Datenbank ist, desto wahrscheinlicher ist es jedoch, dass man einen Datensatz einer Person auch ohne ihren Namen zuordnen kann.

Anonymisierung & Pseudonymisierung sind eine Frage der Merkmale

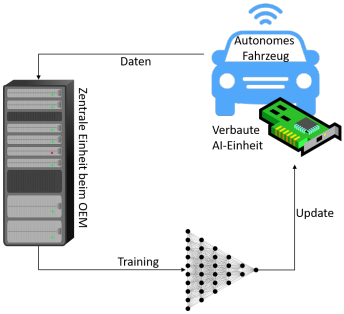
Nehmen wir mich und eine Datenbank aller Einwohner der Stadt (50 000 Einwohner), in der ich wohne. Das Geschlecht halbiert die Zahl. Nun folgt der Bildungsabschluss: Die Promotionsquote eines Jahrgangs liegt um die 2%. Also sind noch ca. 500 Personen übrig ...

Subsidiarität und Autonomie

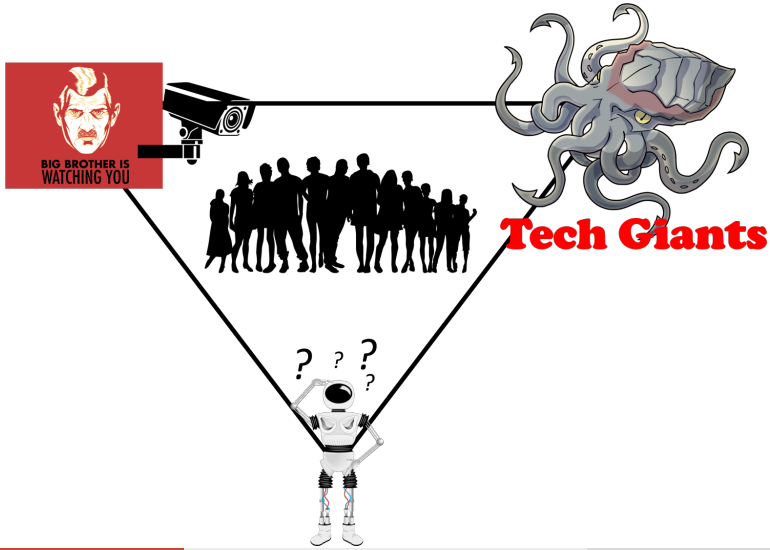
Subsidiaritätsprinzip

Die kleinste Einheit, die dazu in der Lage ist, übernimmt die Aufgaben. Effekt ist eine Stärkung der Selbstbestimmung und Eigenverantwortung dieser Einheit.

- Wem gehören die Daten?
- Wem gegenüber sollte eine KI loyal sein?
- Wie autonom sollten Agenten lernen/arbeiten können?
- Insbesondere wie umgeht man Probleme mit Netzverbindungen oder Vendor lock-in?
- Wer übernimmt Verantwortung?



Subsidiarität, Autonomie und Datenschutz



Robustheit und Sicherheit

- Daten können im Einsatz u.a. verrauscht, manipuliert oder ausgefallen sein
- Gründe sind schwierige Umgebungen, beschädigte Sensoren, absichtliche Manipulationen

Robustheit

Wir nennen ein maschinelles Lernverfahren robust, wenn gilt: *kleine Fehler, kleine Auswirkungen*. Ein Verfahren ist nicht robust, wenn geringe Störungen der Eingangsdaten stark abweichende Ergebnisse liefern.

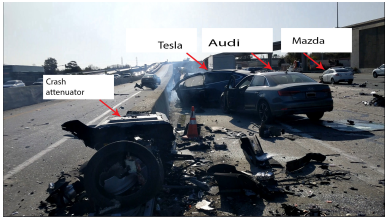
- Die Definition ist verwandt mit der Definition der mathematischen Stabilität eines Problems (gut oder schlecht gestelltes Problem).

Sicherheit

Voraussichtlich störungs- und gefahrenfreie Funktion eines technischen Systems. Sicherheit ist dabei nicht absolut, sondern i.A. wird ein Grad von Unsicherheit akzeptiert.

Robustheit und Sicherheit

- Der Straßenverkehr ist eine komplexe Situation, bei der oft nicht alle Faktoren beobachtbar sind.
- Menschliche Fahrer verursachten 2019 ca. 300 000 Straßenverkehrsunfälle mit Personenschaden, ca. 3000 davon tödlich.
- Was sollte man realistischerweise von autonomen Fahrzeugen erwarten?

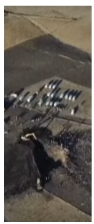


Tesla Crash am 23.03.2018 in Kalifornien (USA)
Quelle [Public Domain]:
`https://commons.wikimedia.org/wiki/File:Mtn_view_tesla_scene_graphic_(28773524958).jpg`

- Laut einer Umfrage im Auftrag des *Verbands der TÜV e. V.* vom Januar 2020 erwarten 40% der Befragten 100% Fehlerfreiheit von einem KI-System (Anteil steigt bei sicherheitskritischen Systemen).
- Das ist illusorisch. Um ein realistischeres Gefühl zu bekommen, werfen wir einen Blick auf einen sehr durchregulierten Sektor, die Luftfahrt.

Robustheit und Sicherheit

- Der Verlust eines Flugzeuges soll gemäß dem Sicherheitskonzept *Extremely improbable* sein
- *Extremely improbable* $= 10^{-9}$ Fehler pro Flugstunde
- Was wäre, wenn es bei einer AI für ein poluläres Auto möglich wäre, diese Qualität zu erreichen?



Boeing 737MAX Ethiopian Airlines
Quelle: `https://youtu.be/PlaMQBEg-9M`
36C3 Talk von Bernd Sieker Folie 6 von 38

- Gehen wir von ca 2,5 Mio. Autos eines Typs aus, die jeden Tag 2 Stunden bewegt werden.
- Gehen wir mal von 365 Tagen pro Jahr und 10 Jahren Einsatzzeit aus.
- *Extremely improbable*:

$$\text{Für das einzelne Auto: } 7300 \text{ Einsatzstunden} \cdot 10^{-9} \frac{\text{Auftreten des Fehlers}}{\text{Einsatzstunden}} = 7.3 \cdot 10^{-6}$$

$$\text{Für die Flotte : } 2\,500\,000 \cdot 7300 \text{ Einsatzstunden} \cdot 10^{-9} \frac{\text{Auftreten des Fehlers}}{\text{Einsatzstunden}} = 18.25$$