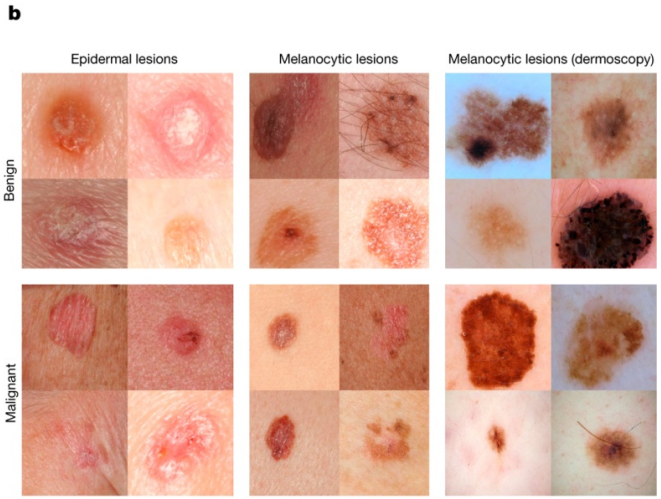
Ethik und KI: Was kann die Technik (nicht) leisten?

Ulrike von Luxburg

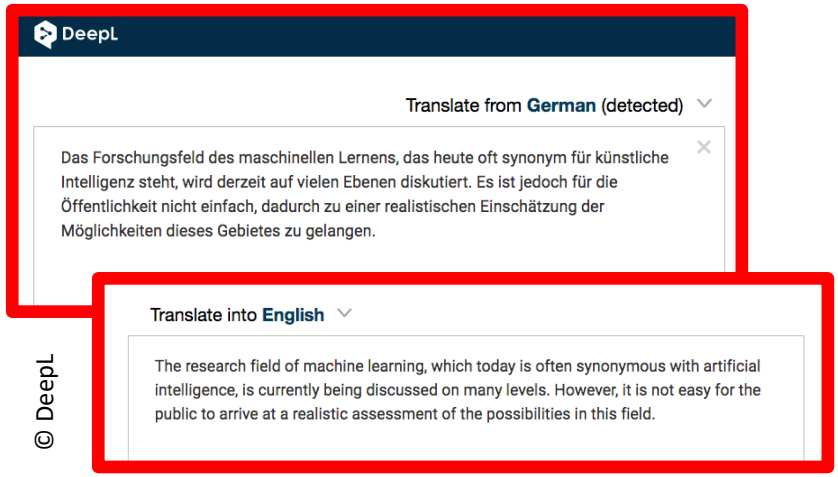
Februar 2021

Anwendungen des maschinellen Lernens



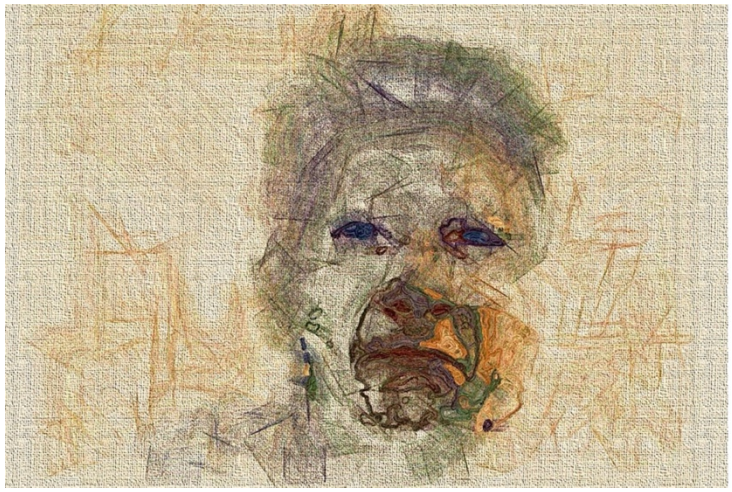
© Esteva et al, Nature 2017

Diagnose von Hautkrebs



© DeepL

Automatische Übersetzung



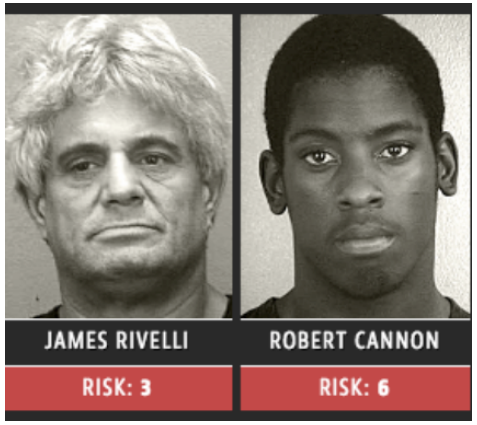
©Simon Colton, University of London

Künstliche Kunstwerke

… aber auch …

©Rekke/Wikipedia

Autonome Waffen



©ProPublica

Vorausschauende Polizeiarbeit

©Dirk Ingo Franke/Wikipedia

Überwachung

… und ein weiter Bereich dazwischen

- Beurteilung von Kreditanträgen
- Vorauswahl von Bewerbungen
- Vergabe von Studienplätzen
- Filtern von Nachrichten
- …

Wie funktioniert KI / ML?

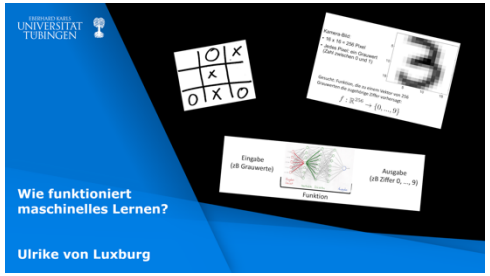
Aufgabe wird definiert, z.B. Hautkrebs erkennen

Trainingsdaten: Bilder von Hautkrebs / gesunder Haut

Lernalgorithmus: durchsucht einen Raum von vorgegebenen Funktionen nach einer, die diese Trainingsdaten gut beschreibt

``Daten + Optimierung + Statistik”

Wie funktioniert maschinelles Lernen?
(links letzte Folie / meine Homepage)



Was ist besonders an ML / KI ?

- Es extrahiert ``Wissen'' aus Daten, die meist nicht für diesen Zweck erhoben worden sind. (-> Bias)

- Es macht Vorhersagen, ohne Modelle zu entwickeln oder Erklärungen zu liefern.
  (-> Erklärbarkeit, Nachvollziehbarkeit)

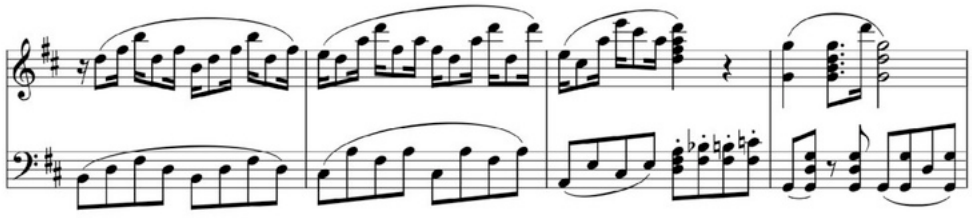
- Riesiges Spektrum an Anwendungen (gute und schlechte)

- ``Jeder'' kann es! (-> Regulierung, Verantwortung)

Intelligenz? Starke KI? Sehe ich nicht!

Derzeitige Durchbrüche beruhen auf maschinellem Lernen. Dahinter stecken Statistik und mathematische Optimierung.

Computer ``suchen“ anders als Menschen, daher sind wir manchmal verblüfft über das Ergebnis. Dann schreiben wir dem Verfahren Kreativität oder Intelligenz zu.



© Avia

© Wikipedia

Meine Meinung: Intelligenz ist derzeit nicht vorhanden und auch nicht am Horizont erkennbar.

Es gibt so viele drängendere Probleme im Bereich Ethik und KI!!! Lassen Sie uns damit anfangen!

Grundproblem: Bias von KI-Systemen

KI-Systeme spiegeln wider, was in den Daten steckt:

Beispiele:

- Gesichtserkennung: funktioniert besser bei weißen als bei schwarzen Menschen
  - Wurde trainiert auf Bildern von überwiegend weißen Menschen

- Unterstützung der Sachbearbeiter*innen im Arbeitsamt durch Klassifizierung von Arbeitslosen

Bias von KI-Systemen - technische Sicht

Bias kann manchmal … teilweise …. korrigiert werden.

Es gibt keine technische Lösung, die Biases automatisch verhindern!

Unbedarfte Anwendungen von ML sind oft problematisch.

KI ist nie ``neutral''.

Grundproblem: Fairness

Bewertung von Kredit-Anträgen:

Bei gleichen Voraussetzungen erhalten Minderheiten weniger Kredite

Fairness – technische Sicht

Was heißt ``fair“?

Es gibt viele sinnvolle Definitionen von Fairness. Sie schließen sich gegenseitig aus, beweisbar!

Wir können Algorithmen bezüglich einer Definition etwas fairer machen, aber nicht für alle gleichzeitig.

Fairness geht immer auf Kosten der ``accuracy“; welchen ``Preis“ sind wir bereit zu zahlen?

Videos zum Thema
(link siehe letzte Folie)

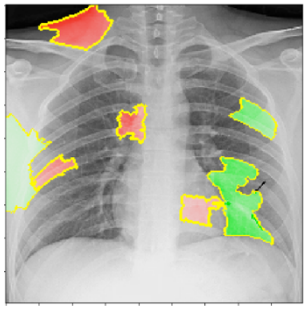
Grundproblem: Erklärbarkeit

Maschinelles Lernen ist intransparent (``black box“).

- Medizinische Anwendungen: Ärzte wollen Erklärungen
- Kreditvergabe: Kunden wollen Erklärung
- Selbstfahrende Autos: Gerichte brauchen Erklärung
- ...

Erklärbarkeit - technische Sicht

Es gibt viele technische Ansätze für Erklärungen:



© COVID-19 Image Data Collection;
Blake VanBerlo/Matt Ross: Investigation of Explainable Predictions of COVID-19 Infection from Chest X-rays with Machine Learning

Aber:

Erklärungen sind Vereinfachungen und führen zwangsläufig einen zusätzlichen Bias ein

Auch automatische Erklärungen können manipuliert werden

Erklärungen können im besten Falle unterstützen, aber wir können uns nicht auf sie verlassen.

Zusammenfassung – technische Sicht

- KI kann nicht perfekt werden!

- Nicht alles, was ethisch wünschenswert ist, ist technisch umsetzbar. (Nicht weil wir es noch nicht können, sondern weil es prinzipiell nicht gehen kann).

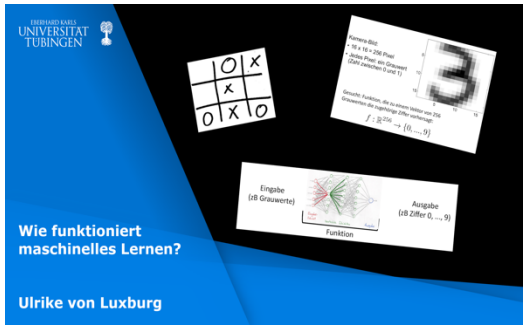
- Brauchen Diskussion, wo KI sinnvoll ist und wo nicht, unter den gegebenen technischen Beschränkungen.

- Brauchen Regulierung und viel mehr Transparenz.

Wenn Sie mehr wissen wollen

Wie funktioniert maschinelles Lernen?
Eine Einführung für jedermann und –frau

Video: https://youtu.be/4QsZkPhNA-A
Text: https://tinyurl.com/ypfj3jse



Vorlesungen “Machine learning in the context of society”

https://youtu.be/5CH2qcZQrpk

Zusammenfassung – persönliche Sicht

- Starke KI macht mir keine Sorgen
- Es gibt Anwendungen von KI, die ich sehr positiv sehe
- Es gibt Anwendungen, die ich sehr kritisch sehe, wo aber klar ist, was zu diskutieren ist (Gesichtserkennung, Überwachung, Waffen).
- Am meisten Unsicherheit herrscht in dem großen Bereich dazwischen:

  Staatliche Organisationen und Firmen, die KI

  - teils unbedarft, teils bewusst -

  für alles Mögliche einsetzen und dadurch Einfluss auf die Gesellschaft ausüben.

Wir brauchen dringend Regulierung und Transparenz!

Grundproblem: Bias von KI-Systemen

Automatische Textanalyse reproduziert Stereotype:

- Mann – Chirurg, Frau – Krankenschwester
- Mann – Programmierer, Frau – Hausfrau
- Weiss – Anwalt, Schwarz – Fussballer

- Automatisches Filtern von Bewerbungen bei Amazon bevorzugte männliche Bewerber
  - In der Vergangenheit waren die meisten ``erfolgreichen'' Angestellten männlich (weil schon immer wenige Frauen eingestellt wurden)