KI – Ein Blick hinter den Hype

Enterprise Architektur Community

Joachim Nelz | S&N Invent

S&N INVENT

Value Driven Performance.

Intro: Hype vs. Realität

Huawei P20:
Embedded KI

Das HUAWEI P20 nutzt die Power von KI, um das Motiv Ihrer Fotos in Echtzeit zu verstehen und die Kameraeinstellungen automatisch anzupassen *<Ref-1>*

"We've made a soft promise to investors that, 'Once we build a generally intelligent system, that basically we will ask it to figure out a way to make an investment return for you.'"

-- Sam Altman, CTO OpenAI (2019)

<Ref-2>

?!?

Full Self Driving no more?!?

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

1	Einführung: Das I der KI
2	Was ist KI / Machine Learning?
3	Herausforderungen
4	Bewertung und Fazit
5	Quellenverzeichnis / Referenzen

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Was ist KI? Eine kritische Betrachtung des Begriffs

- Was ist Intelligenz?
- Woran machen wir das fest?

- Ist KI ein hilfreicher Begriff?

- Was hat KI für Folgen?

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

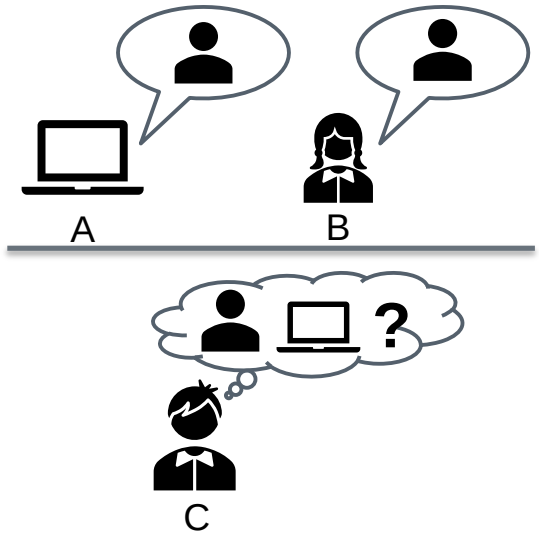
Der Turing Test: Der behaviouristische Ansatz

- Da es keine allgemeingültige Definition des Begriff „Intelligenz“ gibt, statt dessen nur die „Außensicht“, das Verhalten bewerten

Alan Turing,
Meister der Maschine:
"Er knackte im Zweiten Weltkrieg
den Code Enigma "

- I propose to consider the question 'Can machines think?‘ <Ref-3>
- I do not wish to give the impression that I think there is no mystery about consciousness…But I do not think these mysteries necessarily need to be solved before we can answer the question with which we are concerned in this paper. <Ref-3>

-- Alan Turing



© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Der ELIZA Effekt

- "I had not realized ...
  that extremely short exposures
  to a relatively simple computer program
  could induce **powerful delusional thinking**
  in quite normal people." <Ref-4>

  -- Joseph Weizenbaum

Joseph Weizenbaum
(emer. Prof. der Informatik. des MIT).
Aufgenommen: Berlin, Deutschland. 11.2.2005
Creative Commons Copyright:
Ulrich Hansen, GER (Journalist)

- Wir (als Menschen) sind keine objektiven, neutralen Beobachter
  - Wir sind außerordentlich leicht zu täuschen;
  - Wir tendieren dazu, das Objekt zu „vermenschlichen“ d.h. menschliches Verhalten hineinzuinterpretieren

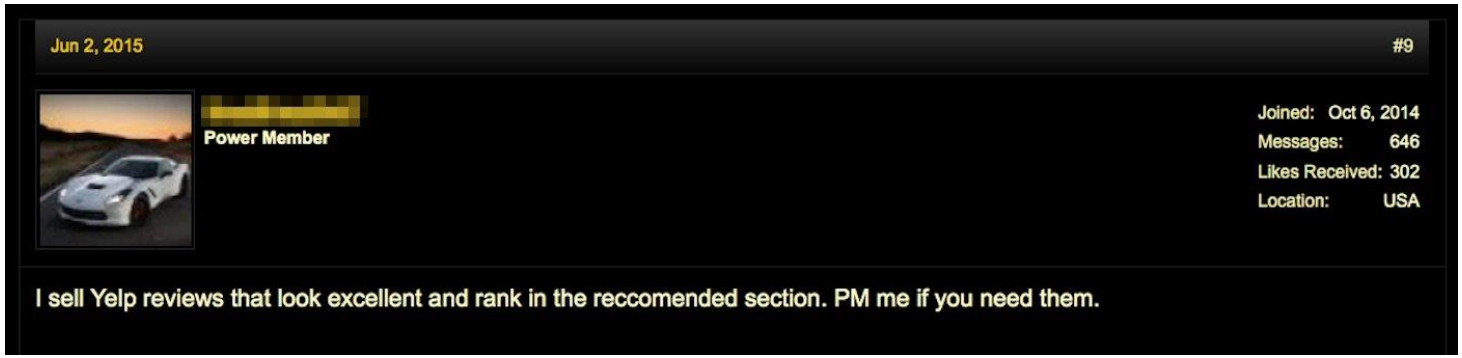
© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Fake Online Reviews durch Social Engineering

- Auf Basis verfügbarer Reviews wurde ein RNN trainiert (2017)
  - We [carried] out a user study (=600) and [showed] that not only can these fake reviews consistently avoid detection by real users, but they provide the same level of user-perceived 'usefulness' as real reviews written by humans.

- I want people to pay attention to this type of attack vector as very real an immediate threat

-- Ben Y. Zhao
<Ref-5>



© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Moravecs Paradox
Einfaches ist schwer, Schweres ist einfach

- It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility”

  -- Hans Moravec, Mind Children

- *Ist also eine komplexe analytische Aufgabe (wie Go) beweiskräftig und geeignet um “Intelligenz in der realen Welt” herzuleiten?*

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

1	Einführung: Das I der KI
2	Was ist KI / Machine Learning?
3	Herausforderungen
4	Bewertung und Fazit
5	Quellenverzeichnis / Referenzen

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Was ist Machine Learning? – Eine Definition

- A computer program is said to learn from experience E with respect to task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

  -- Tom Mitchell, 1997

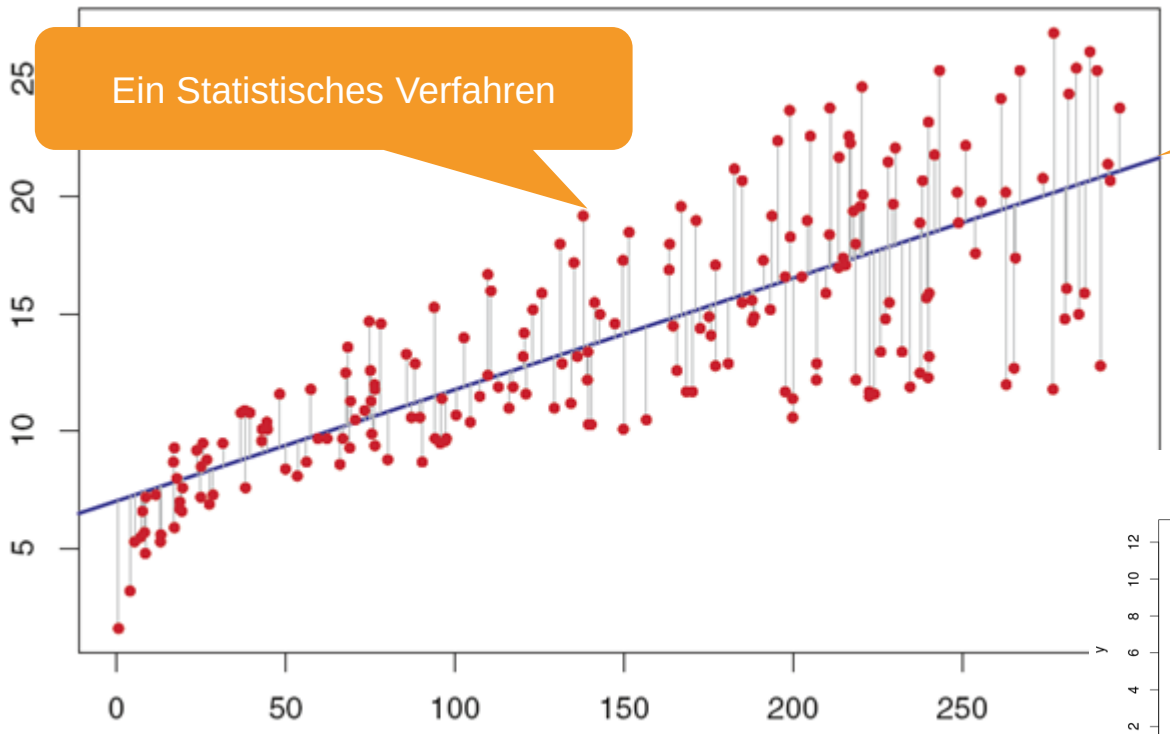
- **Automatische** Verbesserung ohne menschlichen Eingriff
- Fokussiert auf einen spezifischen **Task**
- Erforderlich ist eine **Bewertungsfunktion** (P)

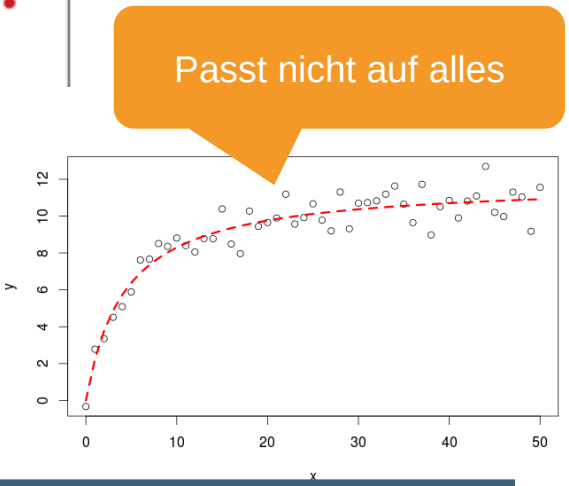
- *Ist Lernen nicht ein Zeichen von Intelligenz?*

- ***Nein!***

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Beispiel: Lineare Regression





Automatisiert, auf einen **Task** abgestimmt, **Bewertung** anhand Minimierung der Abweichung
➔ Ein Machine Learning Verfahren

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Klassifikation von Machine Learning Verfahren

Lernprozess und Feedback

- Supervised, unsupervised, semi-supervised
- Reinforcement Learning

Bildung der Vorhersagen

- Instance-based vs. model-based learning

Zeitpunkt des Trainings

- Online vs. Batch learning

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Supervised Learning

- Trainingsdaten bekommen von Menschen **Labels**

- Typische Tasks
  - Klassifikation: Spam vs. Non-Spam
  - Regression: Numerische Voraussagen AKA „Predictions“

- Typische Algorithmen
  - Linear Regression
  - Support-Vector Machines
  - Neural Networks

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Unsupervised Learning

- Trainingsdaten ohne Labels
  → Lernen ohne menschliche Trainer

- Typische Tasks
  - Segmentierung von Usern / Kunden
  - Zusammenhänge in Daten finden und verstehen
  - Daten vereinfachen

- Typische Algorithmen
  - Clustering
  - Association rule learning
  - Visualization algorithms
  - Dimensionality reduction

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Semisupervised Learning

- Nur teilweise gelabelte Trainingsdaten
- Typische Tasks
  - Personenerkennung z.B. in Google Photos
- Typische Algorithmen
  - Deep Belief Networks
    - Stacked Restricted Boltzmann Machines
  - Kombination von Clustering mit Supervised Learning Ansatz
    - Pro Cluster reicht ein Label aus

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Reinforcement Learning

- Agent := Learning System
  - beobachtet Umgebung, agiert, bekommt Feedback
  - entwickelt eigenständig seine Strategie (AKA Policy)

- Typische Tasks
  - Schach, Go etc.
  - Autonom navigierende Roboter

- Typische Algorithmen
  - AlphaGo
    - Bei seinem Match gegen den Weltmeister Lee Sedol März 2016 gewann AlphaGo allein durch Anwendung seiner gelernten Policy

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Instance-Based Learning

- Trainingsdaten als „auswendig gelerntes“ Raster
- Vorhersagen (Predictions) werden dieser Daten getroffen

- Typische Tasks
  - Spam Mails anhand der Ähnlichkeit zu bekannten Spam Mails klassifizieren
    - Benötigt wird ein Ähnlichkeitsmaß

- Typische Algorithmen
  - K-Nearest Neighbors

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Model-Based Learning

- Aus Trainingsdaten wird ein Modell abgeleitet
- Vorhersagen (Predictions) werden anhand des Modells getroffen

- Typische Tasks
  - Daten miteinander in Beziehung setzen
    - Z.B. Immobilienpreis und Größe des Wohnraums

- Typische Algorithmen
  - Lineare Regression

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

KI als Sammlung von Verfahren

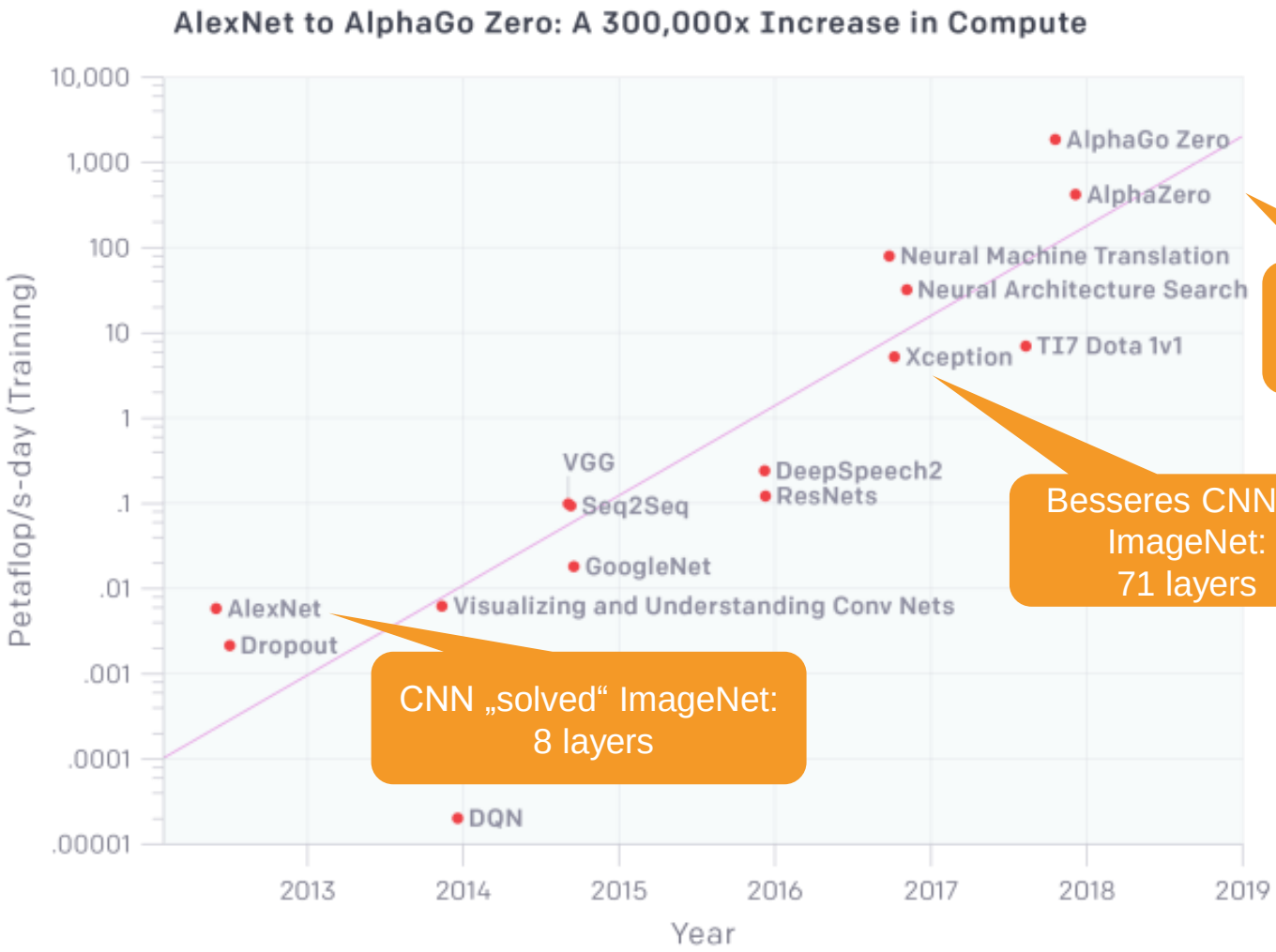
- KI umfasst als Fachgebiet zahlreiche Verfahren
- *Auch Neuronale Netze sind ein Teil davon*

- Es gibt keinen *Über-Algorithmus*, der auf generelle Probleme anwendbar ist

- Für verschiedene Szenarien / Anwendungsfälle kommen spezifische Verfahren in Frage

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

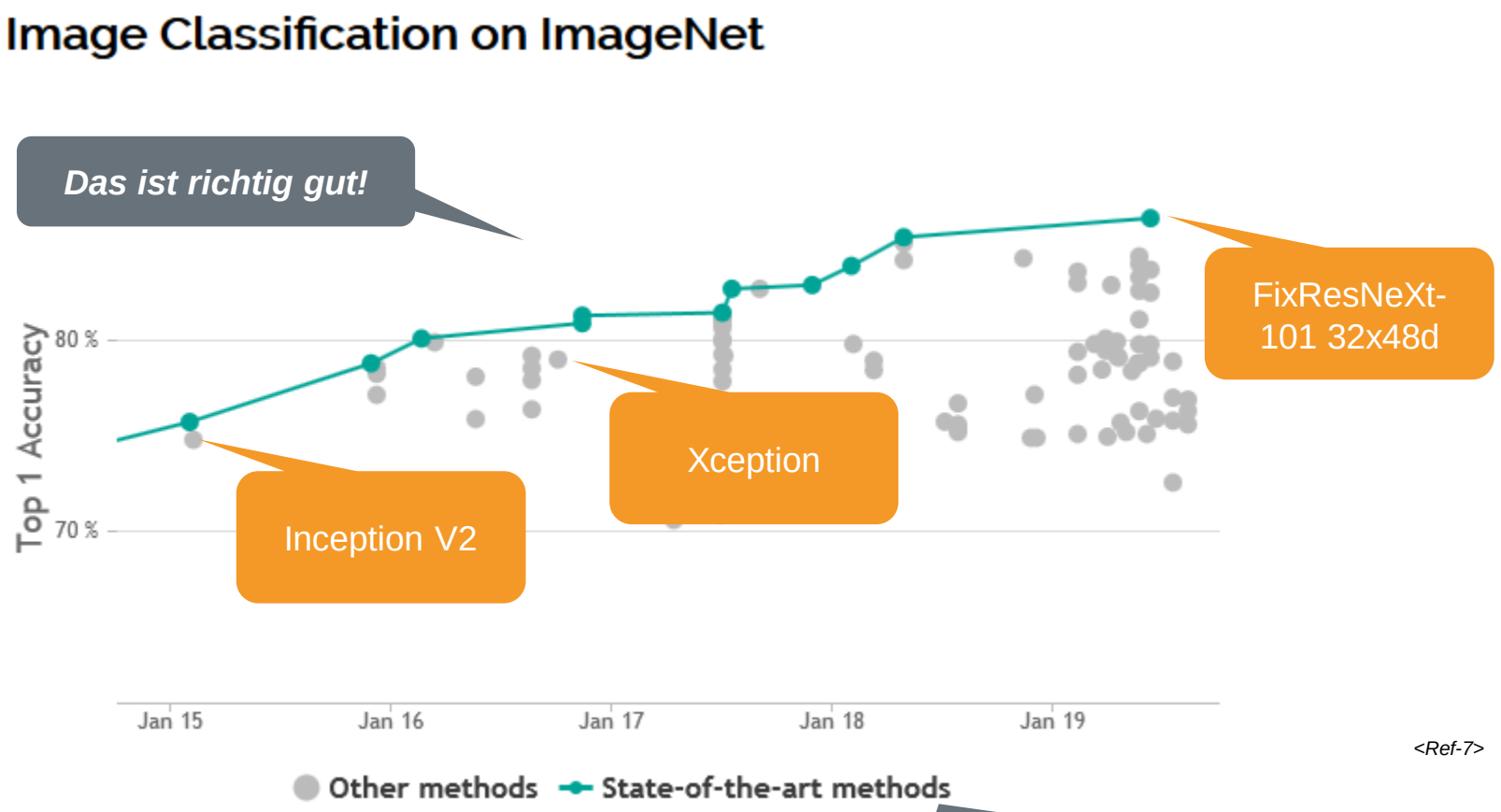
Steigerung der Rechenleistung für komplexere Modelle



<Ref-6>

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Sättigung erreicht?



<Ref-7>

Geht es prinzipiell nicht besser?
Oder brauchen wir andere Verfahren?

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Deep Fail – Computer Vision ist noch nicht gelöst

Image: Imgur

<*Ref-8*>

- Beeindruckende Fortschritte bedeuten noch nicht, dass Computer Vision gelöst wäre.
- Die FaceSwap KI *hat nicht verstanden*, was ein Gesicht ist.

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

1	Einführung: Das I der KI
2	Was ist KI / Machine Learning?
3	Herausforderungen
4	Bewertung und Fazit
5	Quellenverzeichnis / Referenzen

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

No Free Lunch Theorem

- *No reason to prefer one model over another*
  - *„The Lack of A Priori Distinctions Between Learning Algorithms“*

-- David Wolpert, 1996

- Für manche Probleme ist Lineare Regression das beste Modell, für andere Neural Networks…
- **Aber man weiß es nicht vorher**

- Der einzige sichere Ansatz wäre, *alle zu versuchen…*
  - Da das nicht praktikabel ist, hilft gewonnene Erfahrung anhand ähnlicher Daten bei der Einengung auf wenige potentielle Kandidaten.

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

The Unreasonable Effectiveness of Data

- *„Data matters more than algorithms“*

  -- Peter Norvig et al., 2009

- *„These results suggest we may want to reconsider the tradeoff between spending time and money on algorithm development versus spending it on corpus development.“*

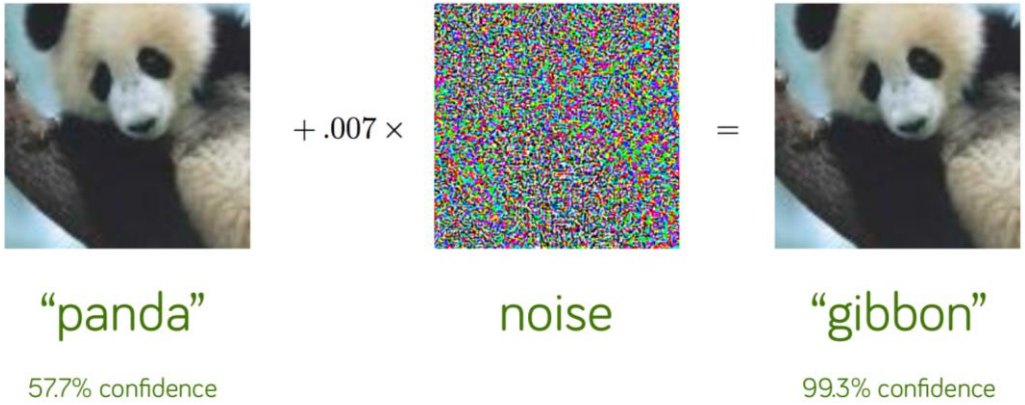
  -- Michele Banko und Eric Brill, 2001

- Aber… nicht immer sind viele Testdaten verfügbar!

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Adversational Attacks on Neural Networks

<Ref-10>, <Ref-13>



<Ref-11>, <Ref-12>

- Seit 2014 ist bekannt: ConvNets sind anfällig für Angriffe *<Ref-13>*
  - Auffällig ist die **hohe Konfidenz der Fehlklassifikation**
  - Nur ein geringe Abwandlung des Inputs genügt
    seit 2017: One pixel attack for fooling deep neural networks *<Ref-9>*

- Verteidigung?
  - *“Many of the most important problems still remain open... We do not yet know whether defending against adversarial examples is a theoretically hopeless endeavor or if an optimal strategy would give the defender an upper ground. “*

-- Ian Goodfellow

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Adversational Attacks – Questionable Robustness <Ref-14>

- Die **Robustheit** der aktuellen Ansätze ist noch gering
  - Das hat Folgen für sicherheitskritische Anwendungen wie Autonomes Fahren (d.h. insbesondere volle Autonomie)

- ***Es handelt sich nicht um „Bilderkennung“*** im Sinn des Verstehens, was Inhalt und Bedeutung der Szene sind
  - Statt dessen: Klassifikation passend zu den Trainingsdaten

Robust Physical-World Attacks on Deep Learning Visual Classification <Ref-14>

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

1	Einführung: Das I der KI
2	Was ist KI / Machine Learning?
3	Herausforderungen
4	Bewertung und Fazit
5	Quellenverzeichnis / Referenzen

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Wo wir stehen – noch am Anfang

- Von „echter Intelligenz“ (was immer das ist) ist man aktuell noch sehr weit entfernt.
- Es handelt sich um spezifische Verfahren, keine „generelle KI“
- Es ist ein langer Weg vom „Lane Assist“ zum „Autopilot“

- Hat also „KI“ noch keine Relevanz?

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Wo wir stehen – bereits mitten in der Transformation

- Neu ist die Fähigkeit, gigantische Datenmengen automatisch zu verarbeiten
- Automatisierte Überwachung von Sensordaten und Kommunikation wurde möglich ***und ist bereits Realität***

- Nicht nur die Technologien, sondern auch **der Begriff KI** hat Auswirkungen

- ***Dies wird nicht nur das Business, sondern die Gesellschaft und das Leben jeden einzelnen verändern***

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Automatisierte Überwachung und Auswertung

- I think we're at high noon in the information age. I say this because of the following. It is really very nearly in our grasp to be able to compute on all human-generated information. You know what's nice about humans compared to sensors? You can only do so much stuff in 24 hours.

  -- Ira Hunt, CTO of the Central Intelligence Agency (2013)

  <Ref-15>

- Establishing the Algorithmic Warfare Cross-Functional Team (Project Maven) to accelerate DoD's integration of big data and machine learning.

  → *bombing and killing*

  - The objective: turn enormous volume of data into **actionable** intelligence and insights at speed
  - The first task [Analysis of UAV video] in support of the Defeat-ISIS campaign

  -- Project Maven, U.S. DoD (2017)

  <Ref-16>

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

The Algorithmic Warfare Cross-Functional Team Project Maven

DEPUTY SECRETARY OF DEFENSE
1010 DEFENSE PENTAGON
WASHINGTON, DC 20301-1010

Projekt Maven hat geliefert…

MEMORANDUM FOR: SEE DISTRIBUTION

SUBJECT: Establishment of an Algorithmic Warfare Cross-Functional Team (Project Maven)

As numerous studies have made clear, the Department of Defense (DoD) must integrate artificial intelligence and machine learning more effectively across operations to maintain advantages over increasingly capable adversaries and competitors. Although we have taken tentative steps to explore the potential of artificial intelligence, big data, and deep learning, I remain convinced that we need to do much more, and move much faster, across DoD to take advantage of recent and future advances in these critical areas.

Accordingly, I am establishing the Algorithmic Warfare Cross-Functional Team (AWCFT) to accelerate DoD's integration of big data and machine learning. The AWCFT's objective is to turn the enormous volume of data available to DoD into actionable intelligence and insights at speed.

<Ref-17>

<Ref-16>

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Folgen des Begriffs KI

- Übersteigerte Erwartungen
  - The Silver Bullet – die Lösung aller Probleme
  - „Hier geschieht ein Wunder“
- Deja vu - Da waren wir schon einmal – im KI Winter
  - Übersteigerte Erwartungen wurden enttäuscht
  - „KI kann nicht liefern“
- Der Test: Kommt der voll autonome Autopilot, oder nicht?

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Folgen des Begriffs KI : Haftungsfrage der Blackbox

- Hersteller kann Funktion nicht nachvollziehen → nicht haften
- „Die KI“ ist keine Person → kann nicht haften
- **→ Verlagerung des Risikos / der Haftung zum User**

- Kern der Haftung im Straßenverkehr ist die Halterhaftung gemäß § 7 Abs. 1 StVG. Die Halterhaftung ist bereits jetzt eine Gefährdungshaftung. *<Ref-18> - <Ref-21>*
  - Dies bedeutet, dass der Halter des Fahrzeugs selbst dann haftet, wenn ihn kein Verschulden trifft.
  - Der Gesetzgeber sieht diese verschuldensunabhängige Haftung als gerechtfertigt an, da der Halter allein durch das Vorhalten eines Fahrzeugs ein Risiko schafft.

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Folgen des Begriffs KI : Volljährige KI – E-Person

- Idee:
  Eine Entität dann als intelligent bezeichnen,
  wenn sie selbst für ihre Handlungen verantwortlich wäre

- „Geschäftsfähige“ / „Volljährige“ KI

- Wie würden Sanktionen aussehen? Verschrotten?

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Folgen des Begriffs KI : Automatischer Bias

- Automatische Bewertung von Menschen
  z.B. Visa, Strafvollzug, Bewerbung etc.

- Es wird „Intelligenz“ und „Vorurteilsfreiheit“ angenommen
  - „Die Maschine entscheidet besser als ein Mensch“
  - Schon lange Realität: Schufa, Rasterfahndung, etc.

- Aber ist das wirklich so?
- Kann ein simples Modell wirklich die Realität abbilden?

- Ist Transparenz gegeben, wie die Entscheidung zustande gekommen ist?
  - Bias In → Bias Out

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Automatisierte Sozial Prognose

<Ref-25>

- Sechs Jahre Haft für unerlaubtes Benutzen eines fremden Autos und mangelnde Kooperation mit der Polizei: Diese Strafe hatte Eric Loomis einem Algorithmus zu verdanken. Nachdem der Amerikaner 2013 in einen Überfall mit Waffengewalt verwickelt gewesen war, erstellte ein Computerprogramm der Firma Northpointe für ihn eine Sozialprognose – basierend auf dem Lebenslauf des Angeklagten und 137 Fragen.

  -- FAZ.net, 11.06.2019
  <Ref-22>

- **COMPAS**, an acronym for Correctional Offender Management Profiling for Alternative Sanctions, is a case management and decision support tool developed and owned by Northpointe (now Equivant) used by U.S. courts to assess the likelihood of a defendant becoming a recidivist.

  -- Wikipedia
  <Ref-23>, <Ref-24>,

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Fazit: Willkommen in der hype-befreiten Zone

- Ein Frage der Reife im Umgang
  Wissen, was eine Technologie kann – und was nicht

- KI Technologien sind praxisrelevant
  - Sehr spezifische u. wirksame Werkzeuge für konkrete Probleme
  - Fähigkeit zur Analyse riesiger Datenmengen

- KI Technologien haben Limitationen
  - „Generelle“ Intelligenz ist noch weit entfernt
  - Robustheit und Skalierbarkeit sind fraglich

- KI / „Intelligenz“ ist *als Begriff nicht hilfreich*
  - Vorsicht vor (Selbst-)Täuschung → Erwartungen korrigieren

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

1	Einführung: Das I der KI
2	Was ist KI / Machine Learning?
3	Herausforderungen
4	Bewertung und Fazit
5	Quellenverzeichnis / Referenzen

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Referenzen

- *Folie 2:*
  - *<Ref-1>: https://www.connect.de/ratgeber/kuenstliche-intelligenz-ki-maschinelles-lernen-technik-hintergruende-risiken-chancen-3198700-7995.html*
  - *<Ref-2>: https://techcrunch.com/2019/05/18/sam-altmans-leap-of-faith/?guccounter=1&guce_referrer_us=aHR0cHM6Ly9ibG9nLnBpZWtuaWV3c2tpLmluZm8vMjAxOS8wNS8zMC9haS1jaXJjdXMtbWlkLTIwMTktdXBkYXRlLw&guce_referrer_cs=pthfg1Hr4cNej5CCkXf-NQ*
- *Folie 5:*
  - *<Ref 3>:* Turing, Alan (October 1950), "Computing Machinery and Intelligence" (PDF), *Mind*, **LIX** (236): 433–460, doi:10.1093/mind/LIX.236.433
- *Folie 6:*
  - *<Ref 4>: Weizenbaum, Joseph (1976). Computer power and human reason: from judgment to calculation. W. H. Freeman. p. 7.*
- *Folie 7:*
  - *<Ref-5>:* https://www.businessinsider.de/researchers-teach-ai-neural-network-write-fake-reviews-fake-news-2017-8
- *Folie 20:*
  - *<Ref-6>: See OpenAI https://www.mathworks.com/help/deeplearning/ref/xception.html*
- *Folie 21:*
  - *<Ref-7>:* https://paperswithcode.com/sota/image-classification-on-imagenet
- *Folie 22:*
  - *<Ref-8>:* https://www.heise.de/select/ct/2018/24/1542703490220001
    https://www.goodtoknow.co.uk/family/faceswap-baby-parents-109980

© Copyright 2020 • S&N Invent GmbH • All rights reserved.

Referenzen

- *Folie 26:*
  - *<Ref-9>: https://arxiv.org/abs/1710.08864*
  - *<Ref-10>: https://blog.piekniewski.info/2016/08/18/adversarial-red-flag/*
  - *<Ref-11>: https://towardsdatascience.com/breaking-neural-networks-with-adversarial-attacks-f4290a9a45aa*
  - *<Ref-12>: https://openai.com/blog/adversarial-example-research/*
  - *<Ref-13>: https://arxiv.org/abs/1412.6572*
- *Folie 27:*
  - *<Ref-14>: https://arxiv.org/pdf/1707.08945.pdf*
- *Folie 31 / 32:*
  - *<Ref-15>: https://gigaom.com/2013/03/20/even-the-cia-is-struggling-to-deal-with-the-volume-of-real-time-social-data/2/*
  - *<Ref-16>: https://dodcio.defense.gov/Portals/0/Documents/Project%20Maven%20DSD%20Memo%2020170425.pdf*
  - <Ref-17> https://www.extremetech.com/internet/267013-google-employees-are-in-revolt-over-pentagon-collaboration-on-project-maven
- *Folie 34:*
  - *<Ref-18>:* https://www.heise.de/newsticker/meldung/Autonomes-Fahren-Wer-haftet-bei-Verkehrsunfaellen-4286037.html
  - *<Ref-19>:* https://www.welt.de/wirtschaft/article184207482/Autonomes-Fahren-Der-Fahrer-soll-immer-haften-selbst-wenn-er-unschuldig-ist.html
  - *<Ref-20>:* https://www.sueddeutsche.de/auto/autonomes-fahren-unfall-haftung-1.4300220
  - *<Ref-21>:* https://deutschland.taylorwessing.com/de/2017-09/der-verkehrsunfall-der-zukunft-wer-haftet-beim-autonomen-fahren
- *Folie 37:*
  - *<Ref-22>: https://www.faz.net/aktuell/rhein-main/algorithmen-werden-in-amerika-bei-gerichtsprozessen-genutzt-16230589.html*
  - *<Ref-23>: https://en.wikipedia.org/wiki/COMPAS_(software)*
  - *<Ref-24>: http://equivant.wpengine.com/northpointe-suite/*
  - *<Ref-25>: http://nautil.us/issue/55/trust/are-algorithms-building-the-new-infrastructure-of-racism*

© Copyright 2020 • S&N Invent GmbH • All rights reserved.