KI – Technische Grundlagen, Algorithmic Decision Making und Ethik

Ramak Molavi, Digital Rights Lawyer
Jörn Erbguth, Legal Tech Consultant

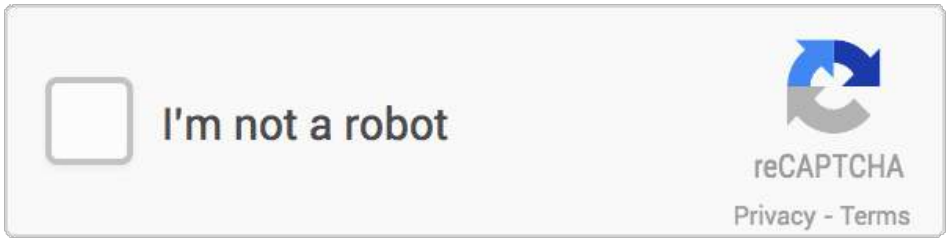
ramak@molavi.de
joern@erbguth.ch +41 787256027

Künstliche Intelligenz

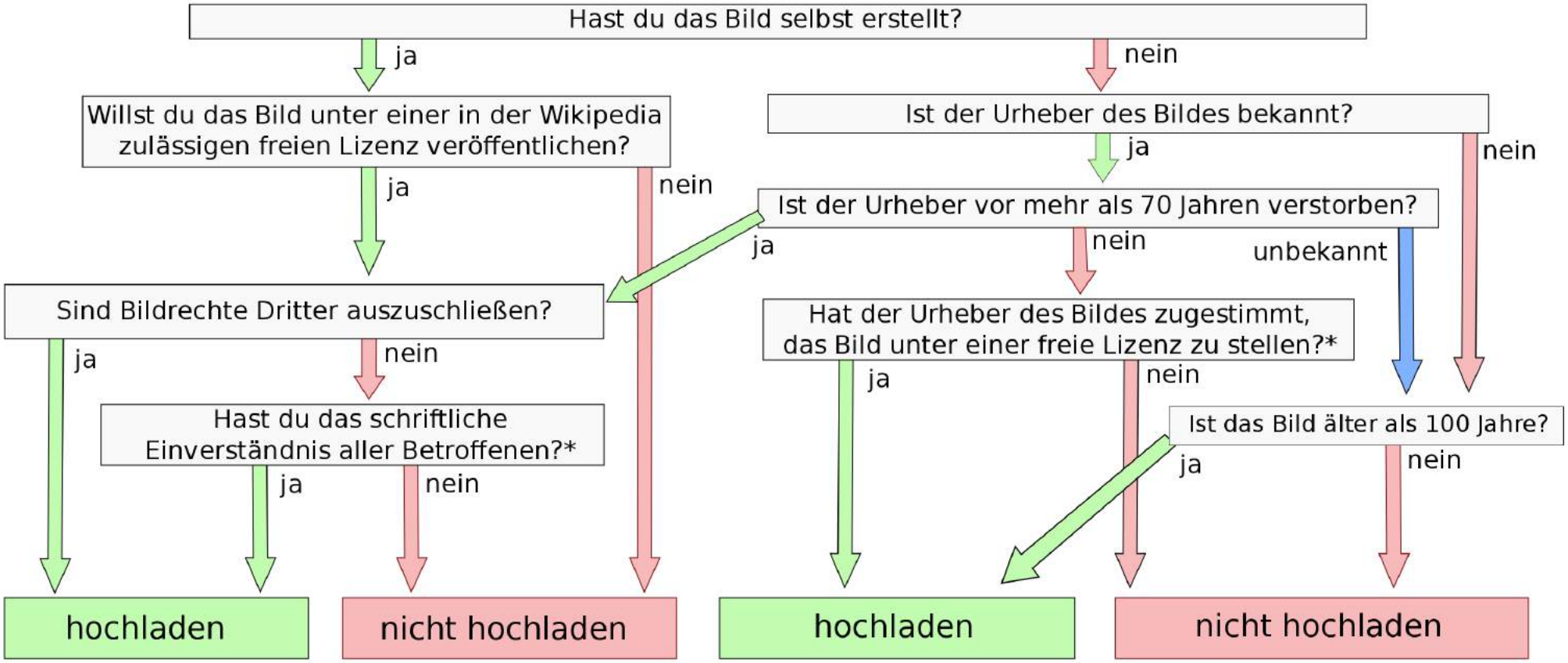
Automatisierung „intelligenten Verhaltens“

Turing-Test


specific
Geben Sie den angezeigten Text ein
reCAPTCHA


I'm not a robot
reCAPTCHA
Privacy - Terms

Klassische Programmierung / Entscheidungsbaum



* Sämtliche Anfragen an den Rechteinhaber sind an das OTRS (permissions-de@wikimedia.org) weiterzuleiten.

Quelle: Wikipedia Bildrechte https://de.wikipedia.org/wiki/Wikipedia:Bildrechte

Regelbasierte Systeme (deduktiv und induktiv)

Kaufpreisforderung § 433 Abs. 2 BGB	
Obersatz	Rechtsfolgen
Kaufvertrag	Kaufpreisanspruch
	Abnahmeverpflichtung

Vertragsschluss §§ 145ff. BGB	
Obersatz	Rechtsfolgen
Angebot	Kaufvertrag
Annahme	

Ausnahme Geschäftsunfähigkeit § 104 BGB	
Obersatz	Rechtsfolgen
Verantwortung des Gläubigers	Gegenleistungsanspruch bleibt bestehen
oder Annahmeverzug	

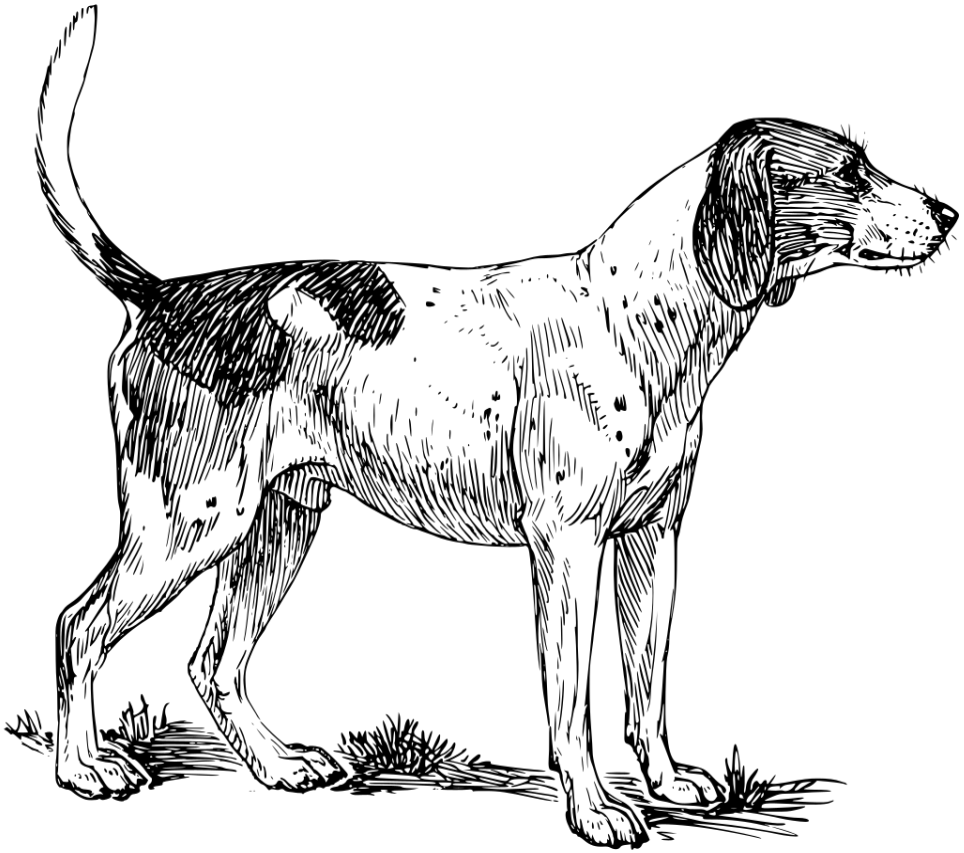
Ausnahme § 326 Abs. 1 BGB	
Obersatz	Rechtsfolgen
Wegfall Leistungspflicht nach § 275 Abs. 1-3	Wegfall des Gegenleistungsanspruchs

Gegenausnahme § 326 Abs. 2 BGB	
Obersatz	Rechtsfolgen
Verantwortung des Gläubigers	Gegenleistungsanspruch bleibt bestehen
oder Annahmeverzug	

Regelbasierte Systeme (deduktiv und induktiv)

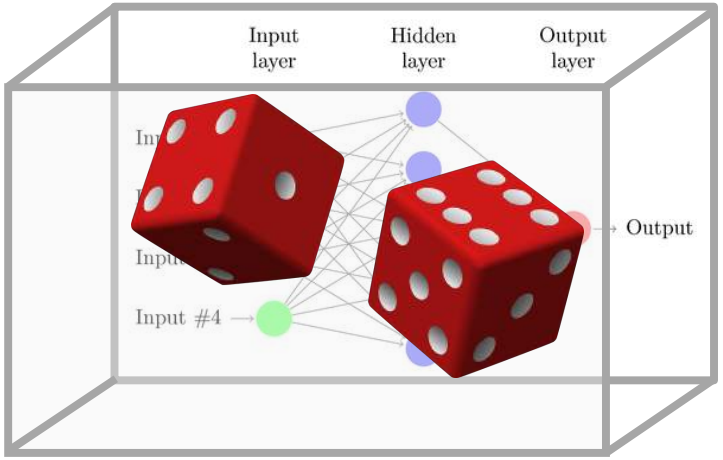
Regelbasierte Systeme (deduktiv und induktiv)

Deep Learning



Wie funktioniert *Deep Learning*? (1)

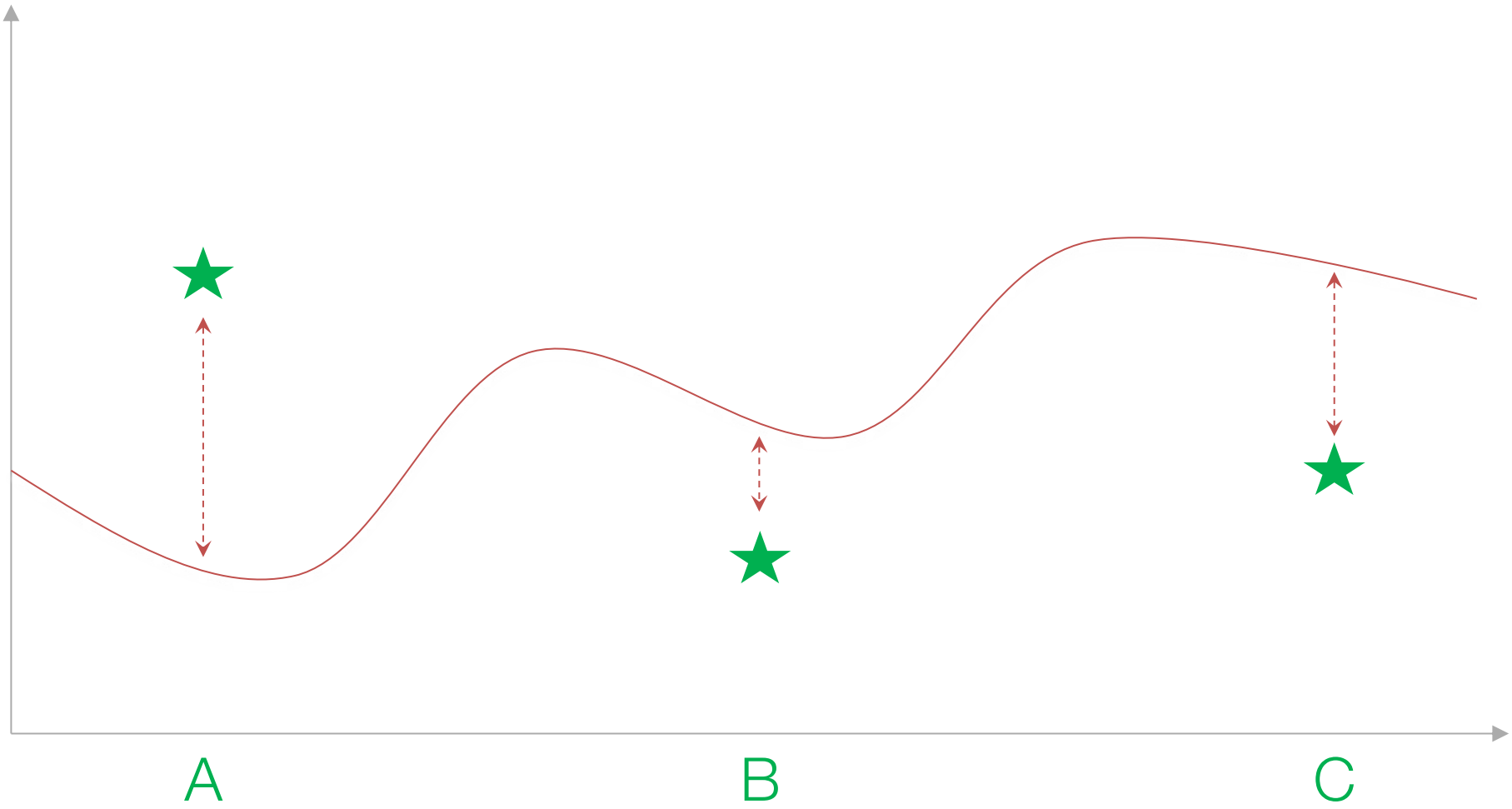
Trainingsdaten



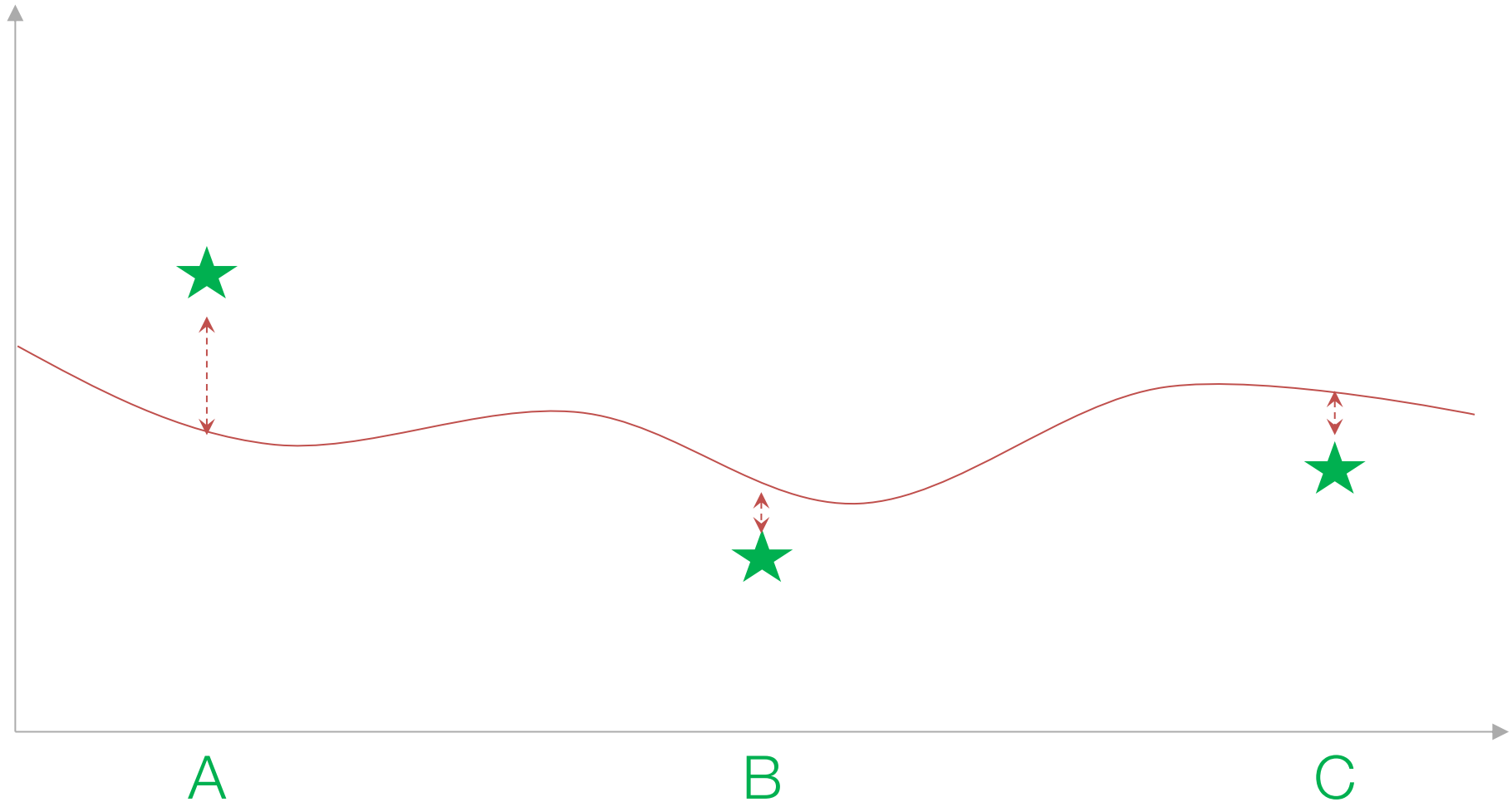
Neuronales Netz

Trainingsalgorithmus

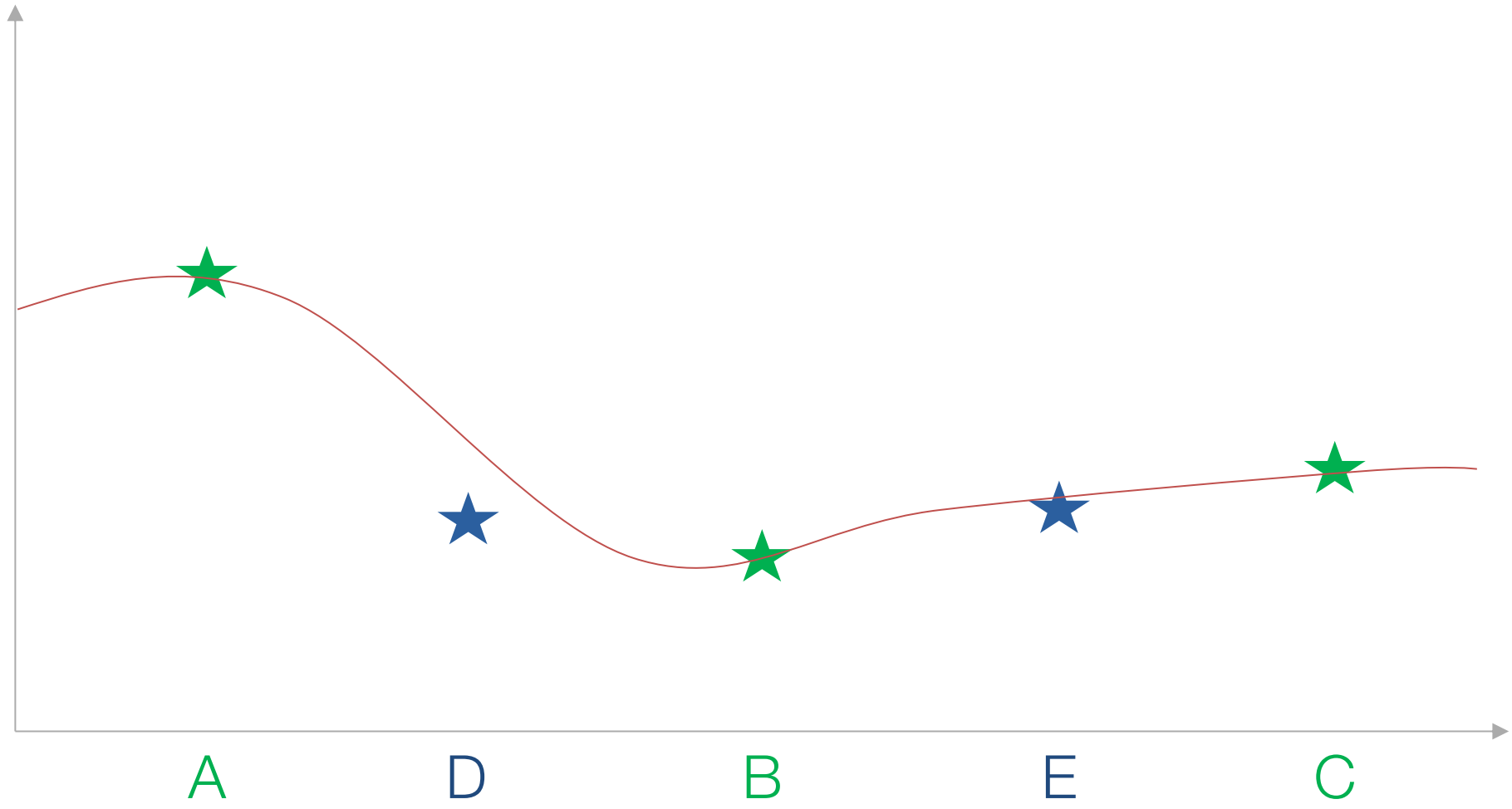
Wie funktioniert *Deep Learning*? (2a)



Wie funktioniert *Deep Learning*? (2b)



Wie funktioniert *Deep Learning*? (2c)



Trainingsdaten

- Trainingsdaten
- Testdaten
- Validierungsdaten

Trainingsdaten

Supervised Learning

Beispiele und Wertungen werden explizit vorgegeben

Unsupervised Learning

Das System bildet selbständig Cluster und erkennt Ausreißer

Reinforcement Learning

Das System probiert aus, oder spielt gegen sich selbst und lernt daraus

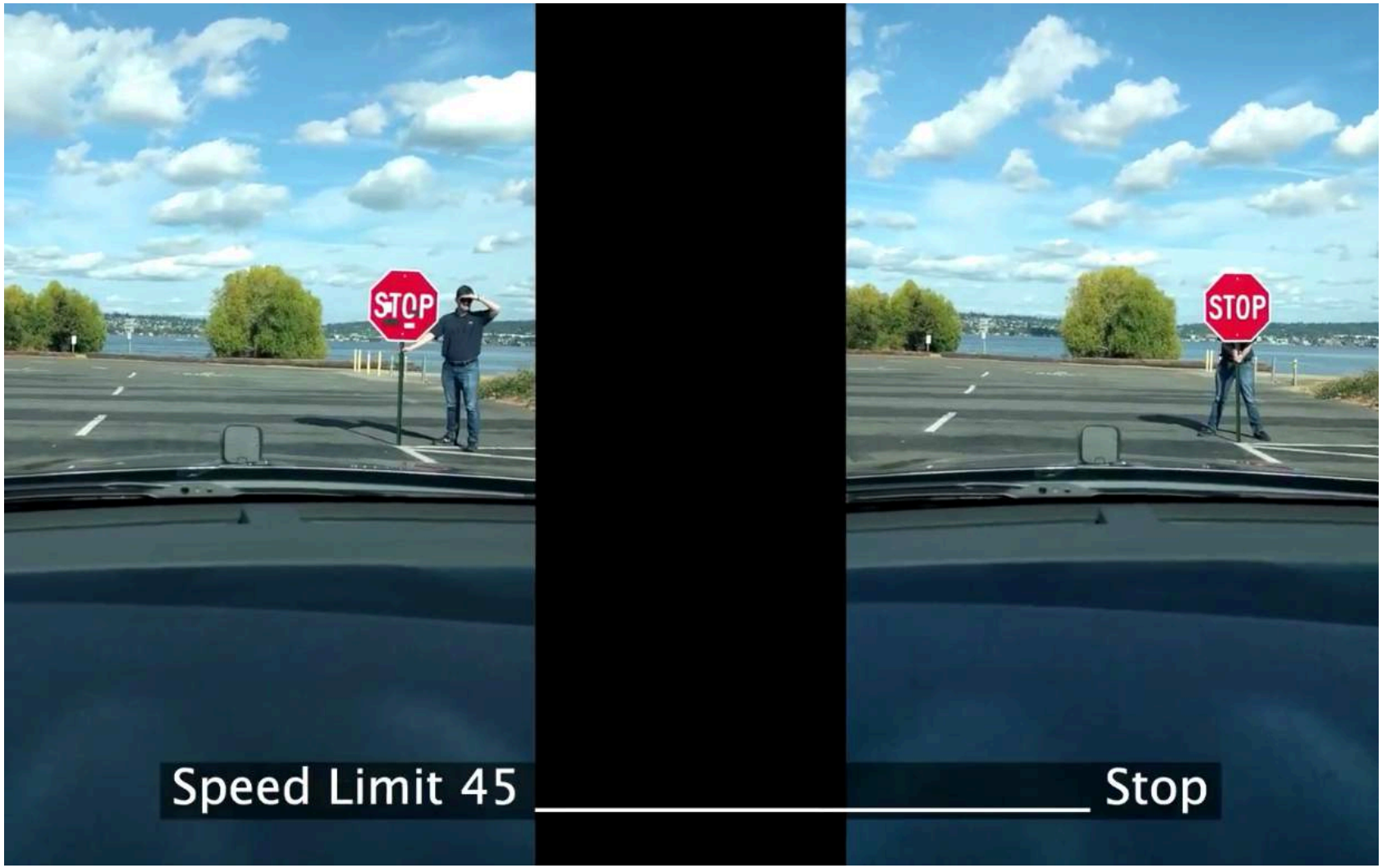
Deep Learning

- Lernt mit Daten und Zufall
- Basiert auf Verallgemeinerung von Zusammenhängen
- Braucht keine Regeln und bekommt diese auch nicht
- Imitiert intelligente Entscheidungen
- Verhalten ist ohne konkrete Tests nicht sicher voraussagbar

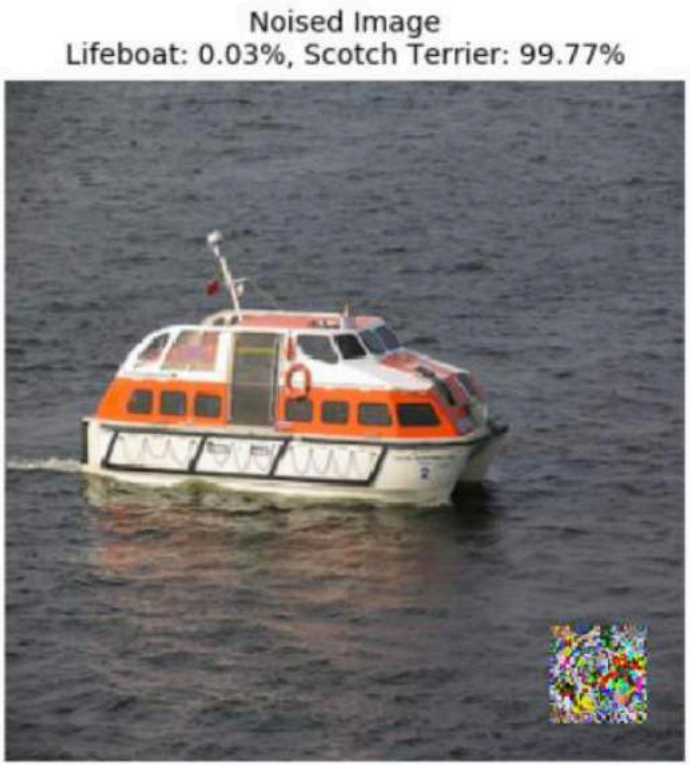
Experiment

Experiment https://teachablemachine.withgoogle.com/
Video https://erbguth.ch/DeepLearning.mp4

Hacking von Deep-Learning (1)



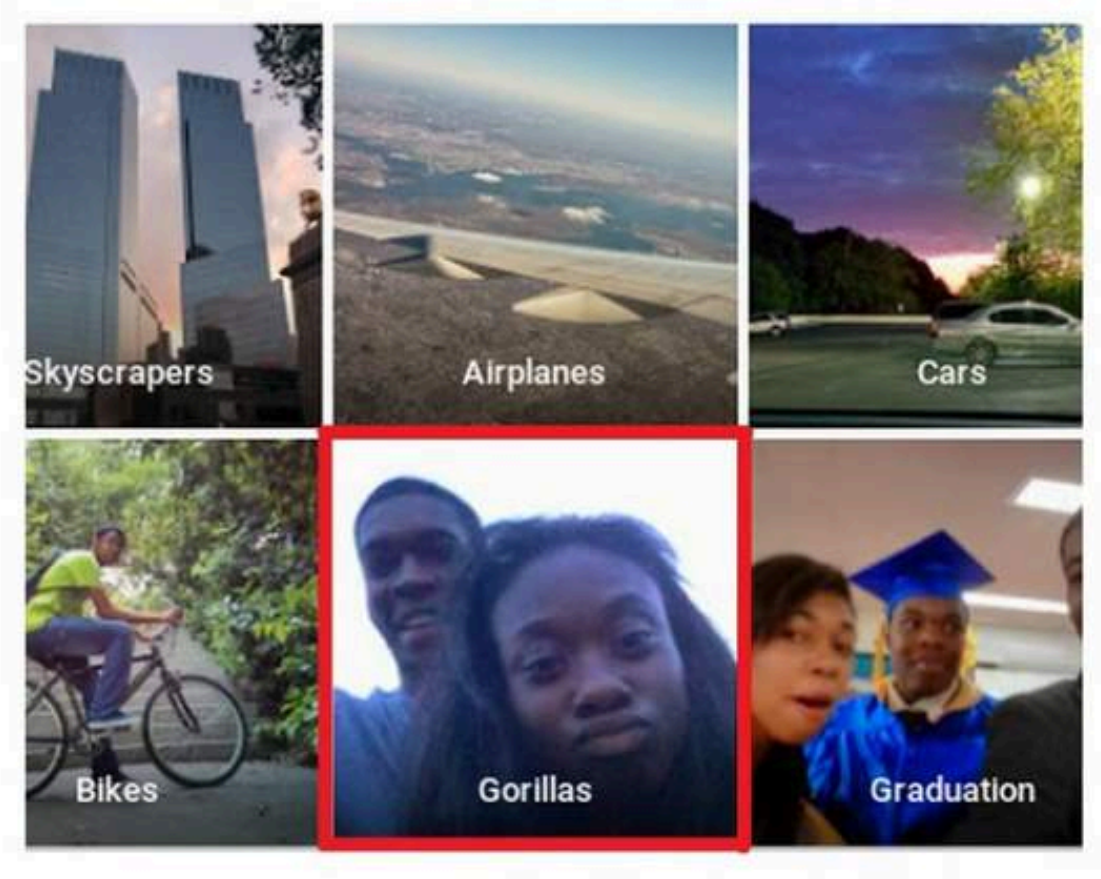
Hacking von Deep-Learning (2)



Lifeboat (89,2%) → Scotch Terrier (99,8%)

Diskriminierung

Lücken in den Trainingsdaten



Woher kommt die Diskriminierung?

- Bias in den Trainingsdaten
- Unvollständige Trainingsdaten
- System nimmt falsche Entscheidungskriterien
- Statistik

Fazit

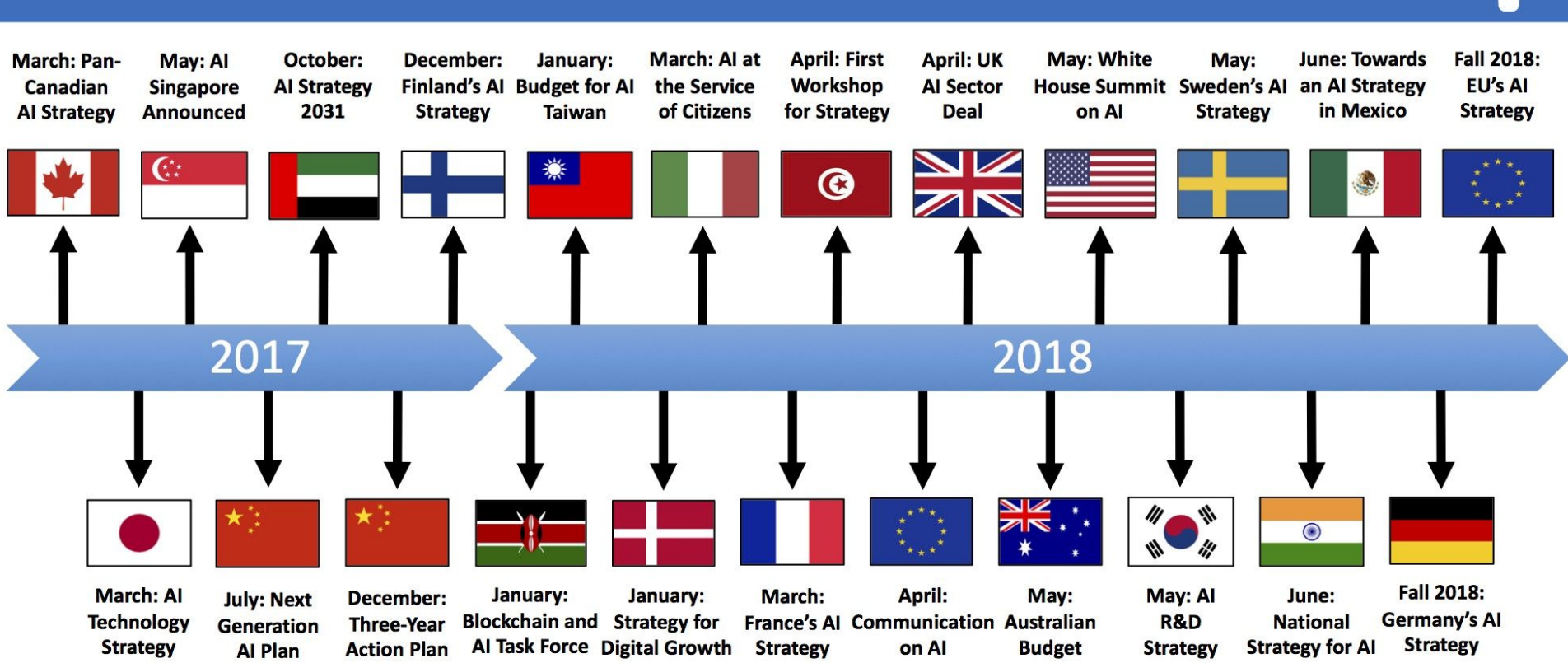
- Deep Learning bietet statistisch gute Ergebnisse
- Deep Learning wird nicht mit Regeln und Werten trainiert
- Das Trainieren von Deep Learning geht nicht ohne Zufall
- Auf Statistik beruhenden Entscheidungen sind problematisch (mit und ohne Deep Learning)

Künstliche Intelligenz

Algorithmische Entscheidungs-findung und Ethik

16. Deutscher IT Rechtstag 2019 - Berlin

Artificial Intelligence Strategies



2018-07-13 | Politics + AI | Tim Dutton

ADM

(algorithmic decision making): algorithmenbasierte Prognose- und Entscheidungsprozesse

- Auswahl von Bewerber
- Zugang zu Schulen
- Prognose von Musik Hits
- Plots von Serien und Filmen
- Medizin, etwa bei der Krebsdiagnose
- Kognitives Kochen etc.
- Auffinden von Mails verärgerter Kunden (Vorauswahl auch durch Chatbots)

- Warenempfehlungen
- präventive Polizeiarbeit
- Optimierung von Energieverbrauch in Industrieanlagen
- Automatisiertes Schreiben Journalistischer Texte
- “Heraussuchen” relevanter Dokumente im Rahmen von Unternehmenskäufen

Ethikdiskussion

Wir Juristen wundern uns.

Soft law?
Ersatz für Recht im KI Bereich?
Was für Philosophen?
Whitewashing?
Berufsethos?

Mythos “unreguliert”

1. Viele Normen gibt es bereits

Nationales Recht:

- GG, AGG
- Begründungspflichten
- Gesetze zum Datenschutz
- Haftungsgrundsätze
- Schutz des Wettbewerbs

EU-Primärrecht

- Verträge der Europäischen Union
- Grundrechte Charta

EU-Sekundärrecht

- DSGVO
- Produkthaftungsrichtlinie
- VO über den freien Verkehr nicht personenbezogener Daten

- Antidiskriminierungsrichtlinien
- Verbraucherrecht
- Richtlinien über Sicherheit und Gesundheitsschutz am Arbeitsplatz
- UN-Menschenrechtsverträge und die Übereinkommen des Europarates (zB. Europäische Menschenrechtskonvention)

Domänenspezifische Regulierung

- Medizinprodukteverordnung

Positives Recht

- EU-Charta Artikel über die "Freiheit, ein Unternehmen zu führen"
- die "Freiheit der Künste und Wissenschaften"

2. Faktische Regelungsmacht von Tech Monopolen:

Anbieter und Entwickler von KI setzen seit Jahren faktische Standards

Wie hilft nun Ethik?

ERKENNTNIS 1:

Eine Regulierung, die NACH der Konstruktion ansetzt/ bedacht wird ist nicht wirkungsstark.

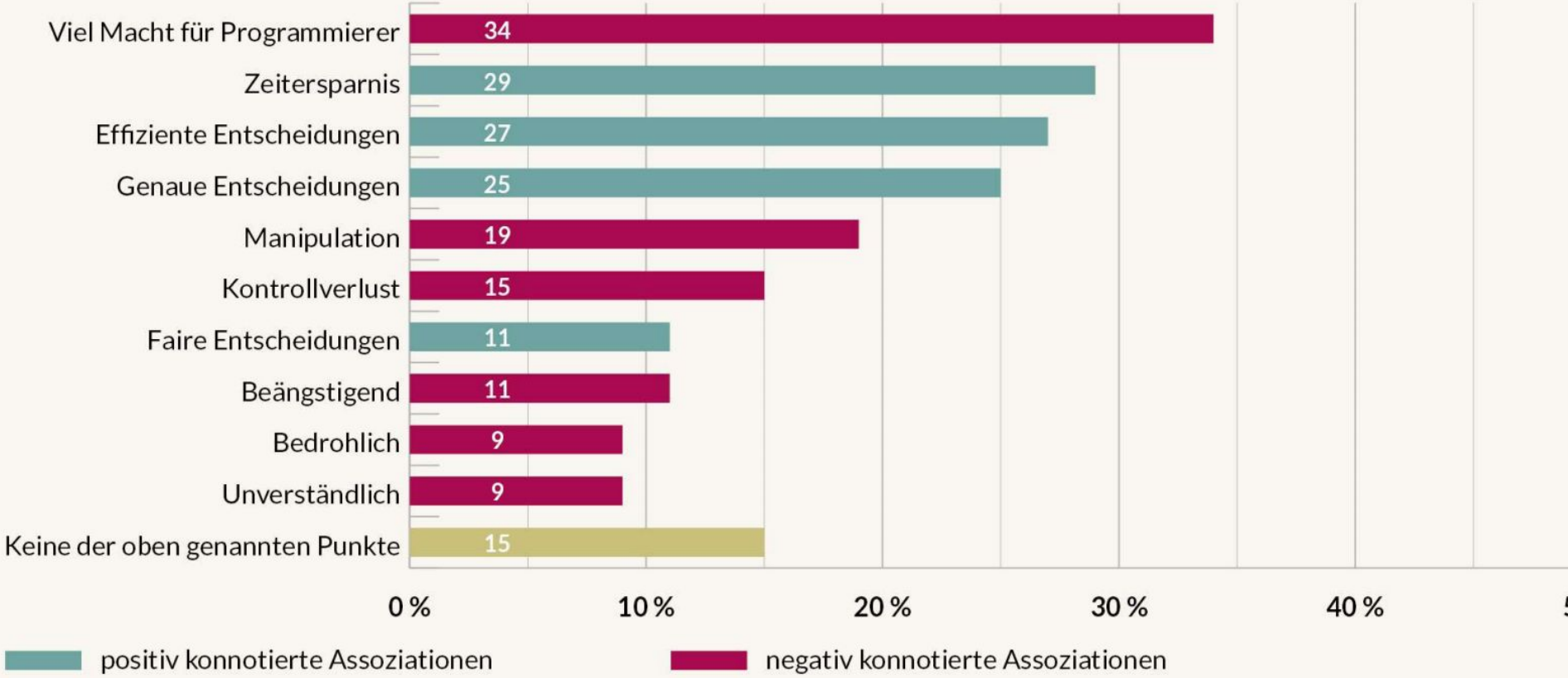
Die Systeme müssen bereits so gebaut werden, dass eine Regulierbarkeit überhaupt möglich ist.

ERKENNTNIS 2:

Die Entwickler besitzen eine Wissenshoheit hinsichtlich der Möglichkeiten ihrer Systeme.
Sie sollten Prinzipien des Gemeinwohl bereits in die DNA einstricken.
So wie es beispielsweise auch mit Sicherheitsaspekten geschieht.

ABBILDUNG 4 **Assoziationen zum Begriff Algorithmus | EU**

„Was kommt Ihnen in den Sinn, wenn Sie das Wort ‚Algorithmus' hören? "



Quelle: Repräsentative Online-Umfrage von Dalia Research im Auftrag der Bertelsmann Stiftung, September 2018, n = 10.960 Befragte

With great power comes great responsibility

Ethik verlagert die Diskussion vor.

Herausforderungen & offene Fragen

Autonome Entscheidungen können eine Bedrohung für den freien Willen bedeuten oder die Übernahme von Verantwortung gefährden.

> Ist ein übermäßiges Vertrauen in vorgeblich „neutrale“ und fehlerlose maschinelle Entscheidungen angemessen?
> Wie reizvoll ist es für uns Menschen, Entscheidungen, Einschätzungen und damit Verantwortung an Maschinen abzugeben?

Diskriminierungsverstärkung und Ausschluss:

KI-Systeme können Tendenzen verstärken und Diskriminierungen mit sich bringen.
Ist das beabsichtigt, kann es schlimme Folgen haben, jedoch noch weit problematischere, wenn solche Tendenzen beim maschinellen Lernen unbeabsichtigt entstehen und dann unbemerkt Menschen diskriminiert oder ausgeschlossen werden.

Algorithmische Profilbildung von Menschen:

Das persönliche Zuschneiden von Angeboten oder Services kann zum Teil sinnvoll sein, andererseits aber diese errechnete Personalisierung eine Bedrohung für gesellschaftliche Werte wie politischen und kulturellen Pluralismus sein. (Filter Bubble, Profiling, Scoring, Koppelung mit Preisdifferenzierung, Dual Use etc.)

Ansammlung riesiger Datenmengen für maschinelles Lernen:

Datenschutzgesetze dafür geschaffen wurden, Individuen den Schutz ihrer persönlichen Daten zu gewähren. Das kann diesen großen Datensammlungen im Weg stehen.

Wie kann eine Balance zwischen diesen konträren Zielen aussehen, was muss neu verhandelt werden?

Herausforderungen bei der Auswahl von Daten in Qualität, Quantität und Relevanz:

> Daten die von Software verarbeitet werden sollen, sind nicht unbedingt korrekt und können auf systematische Fehler beruhen.
> Man sollte Software Ergebnissen gegenüber kritisch bleiben und kein übermäßiges Vertrauen in ADM Systeme entwickeln.

Spannungsverhältnis

Gemeinwohlorientierung
&
Solidarprinzip

Innovation
&
Fortschrittsgeschwindigkeit

Ethische Richtlinien für Vertrauenswürdige KI

8. April 2019

ETHICS GUIDELINES FOR TRUSTWORTHY AI

Vertrauenswürdige KI muss:

1. Gesetzeskonform sein
2. Ethische Grundregeln implementiert haben und beachten
3. Sowohl technisch als auch gesellschaftlich robust sein

Gesetzeskonform, da es bereits Regelungsrahmen auch für KI gibt:

Eine Reihe von rechtsverbindlichen Regeln auf europäischer, nationaler und internationaler Ebene gelten bereits heute oder sind für die Entwicklung, den Einsatz und die Nutzung von KI-Systemen relevant.
Respekt vor der individuellen Freiheit und der Menschenwürde.

FUNDAMENT 1: Ethische Grundsäulen

Respekt der menschlichen Autonomie

Schadensvermeidung

Fairness

Erklärbarkeit

Respekt der menschlichen Autonomie

Die Grundrechte sind darauf ausgerichtet, die Achtung der Freiheit und Autonomie der Menschen zu gewährleisten.

In der Interaktion mit KI-Systemen muss die Selbstbestimmung aufrecht erhalten bleiben und die Teilhabe am demokratischen Prozess gewährleistet sein.

KI-Systeme sollten Menschen:

- Nicht ungerechtfertigterweise unterordnen, zwingen, täuschen, manipulieren oder konditionieren.

- Stattdessen sollten sie darauf ausgerichtet sein, die menschlichen kognitiven, sozialen und kulturellen Fähigkeiten zu erweitern, zu ergänzen und zu stärken.

- Dies bedarf bürgerzentrierte Designprinzipien und sinnvolle Wahlmöglichkeiten für den Menschen.

Schadensvermeidung

KI-Systeme sollten weder Schäden (geistige/körperliche) verursachen noch verschlimmern oder anderweitig den Menschen beeinträchtigen.

- KI-Systeme und die Umgebung, in der sie betrieben werden, müssen sicher und geschützt sein. Sie müssen technisch robust sein und es sollte sichergestellt sein, dass sie nicht für bösartigen Gebrauch zugänglich sind.

Schadensvermeidung

- Besondere Aufmerksamkeit bei:
  - Gefährdete Personen
  - Macht- oder Informationsasymmetrien
    (zB. zwischen Arbeitgebern und Arbeitnehmern, Unternehmen und Verbrauchern oder Regierungen und Bürgern.)

- Berücksichtigung der natürlichen Umwelt und aller Lebewesen.

Fairness

Die inhaltliche Dimension des Begriffs:

- Gewährleistung einer gleichmäßigen und gerechten Verteilung von Nutzen und Kosten
- Sicherstellung einer Freiheit vor Diskriminierung und Stigmatisierung.
- Chancengleichheit beim Zugang zu Bildung, Gütern, Dienstleistungen und Technologien.
- Keine Täuschung oder Beeinträchtigung von Menschen in ihrer Wahlfreiheit.
- Verhältnismäßigkeit zwischen Mitteln und Zielen.

Fairness

Die prozessuale Dimension des Begriffs:

- Fähigkeit, Entscheidungen von KI-Systemen und den sie betreibenden Menschen anzufechten und einen wirksamen Rechtsbehelf einzulegen.

- Hierfür muss die für Entscheidung verantwortliche Stelle identifizierbar sein, und die Entscheidungsprozesse sollten erklärbar sein.

Erklärbarkeit

Entscheidend für das Vertrauens der Benutzer in KI-Systeme.

- Transparente Prozesse, offene Kommunikation von ”Fähigkeiten” und Zweck von KI-Systemen.
- (soweit wie möglich) Erklärbare Entscheidungen gegenüber den direkt und indirekt Betroffenen.
- Eine Erklärung, warum ein Modell eine bestimmte Ausgabe oder Entscheidung erzeugt hat (und welche Kombination von Inputfaktoren dazu beigetragen haben) ist nicht immer möglich.

Erklärbarkeit

Diese Fälle werden als "Black Box"-Algorithmen bezeichnet und erfordern besondere Aufmerksamkeit.

- Mögliche Maßnahmen zur Erklärbarkeit bei Black Box KI sind z.B. Rückverfolgbarkeit, Auditierbarkeit und transparente Kommunikation über die Systemeigenschaften.

- Der Grad der Erforderlichkeit hängt von Kontext und der Schwere der Folgen ab

FUNDAMENT 2: Robustheit

Zur Vermeidung von unbeabsichtigten Schäden und Auswirkungen müssen diese Systeme sollten sicher, geschützt und zuverlässig funktionieren, und es sollten Schutzmaßnahmen vorgesehen werden.

Sonst kann kein Vertrauen erwachsen.

FUNDAMENT 2: Robustheit

Dies ist sowohl aus technischer Sicht als auch aus sozialer Sicht (Kontextes und Umgebung, in der das System betrieben wird) erforderlich.

Ethische und robuste KI sind daher eng miteinander verflochten und ergänzen sich gegenseitig.

REALISIERUNG
7 Schlüsselanforderungen an vertrauenswürdige KI:

1. Menschliches Handeln und Aufsicht
2. Technische Robustheit und Sicherheit
3. Datenschutz und Datenmanagement (Data Governance)
4. Transparenz
5. Vielfalt, Nichtdiskriminierung und Fairness
6. Gesellschaftliches und ökologisches Wohlergehen
7. Verantwortlichkeit

Beispiel ”Menschliches Handeln und Aufsicht”

Menschliches Handeln: Benutzer sollten die Kenntnisse und Werkzeuge erhalten, um KI-Systeme in zufriedenstellendem Maße zu verstehen und mit ihnen zu interagieren und - wenn möglich - das System angemessen zu bewerten oder anzufechten.

KI-Systeme können menschliches Verhalten steuern, da sie unterbewusste Prozesse nutzen können, einschließlich verschiedener Formen unfairer Manipulation, Täuschung, Konditionierung, die die individuelle Autonomie gefährden können.

Beispiel menschliches Handeln und Aufsicht:

Das allgemeine Prinzip der Benutzerautonomie muss im Mittelpunkt der Funktionalität des Systems stehen.

Entscheidend dafür ist das Recht, einer Entscheidung nicht allein auf der Grundlage der automatisierten Verarbeitung zu unterliegen, wenn dies rechtliche Auswirkungen auf die Nutzer hat oder sie in ähnlicher Weise erheblich beeinträchtigt.

Beispiel menschliches Handeln und Aufsicht:

Menschliche Aufsicht: Die menschliche Aufsicht trägt dazu bei, dass ein KI-System die menschliche Autonomie nicht beeinträchtigt oder andere schädliche Auswirkungen hat.
Die Aufsicht kann durch Governance (Steuerungs) -Mechanismen wie (HITL) (HOTL) (HIC) erfolgen.

Dies kann die Entscheidung beinhalten, in einer bestimmten Situation kein KI-System zu verwenden.

Beispiel menschliches Handeln und Aufsicht:

Darüber hinaus muss sichergestellt sein, dass die Durchführung der Aufsicht und die Durchsetzung von Rechten möglich ist und möglich bleibt.

- Guter offizieller Anfang
- Vorbild für andere (Länder)
- Betonung, dass es jetzt um die Umsetzung geht
- Partizipationsprozess geht weiter
- Best Practices werden erbeten

- Trustworthy AI?
- Zu viele relativierende Formulierungen “soweit möglich”
- Nicht bindend
- Rote Linien wurden beauftragt, jedoch nicht als solche übernommen. “Rote Linien” sind nun in “Bedenken” umbenannt

Prof. Dr. Thomas Metzinger und der ML Urs Bergmann hatten die Aufgabe „Red Lines“ zu erarbeiten – nicht-verhandelbare ethische Prinzipien, die festlegen, was in Europa mit KI nicht gemacht werden darf.

- Der Einsatz von tödlichen autonomen Waffensystemen
- Das Verbot einer KI-gestützten Bewertung von Bürgern durch den Staat (Social Scoring)
- Kein Einsatz von KI, die Menschen nicht mehr verstehen und kontrollieren können.

Critical concerns

- Scoring sollte nur mit entsprechender Rechtfertigung erfolgen
- Der Mensch sollte erkennen, wenn er mit einer KI kommuniziert damit er die Möglichkeit hat, dies selbst einzuschätzen.
- Tödliche autonome Waffensysteme (LAWS): es braucht eine dringende Entwicklung einer gemeinsamen, rechtsverbindlichen Position.
- Potenzielle langfristige Bedenken: Ein risikobasierter Ansatz schlägt vor, dass diese Bedenken berücksichtigt werden sollten, und zwar im Hinblick auf mögliche unbekannte Unbekannte und "black swans".

Black Swan*
Zufällige und unerwartete Phänomene, die einen großen oder unverhältnismäßigen Einfluss haben.

(aus der modernen kognitiven Theorie)

Transparenz & Erklärbarkeit

Menschliche Kontrolle

Diskriminierungsfreiheit

Klare Verantwortlichkeit

Selbstbestimmung & Autonomie

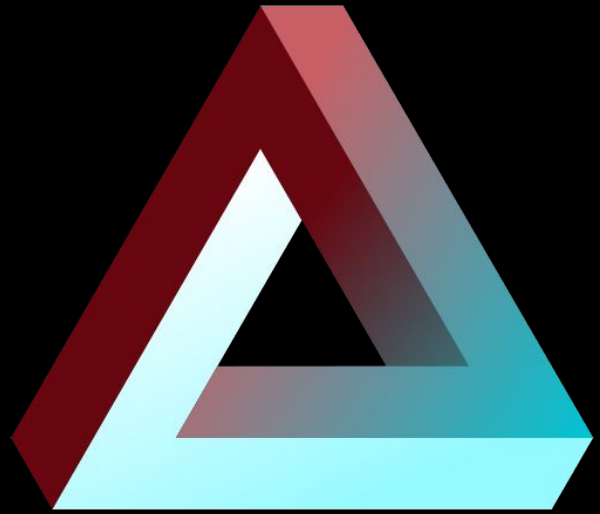
Gemeinwohlzuträgliche Entwicklung von KI

Technik ist nicht neutral

Ramak Molavi

Rechtsanwältin

ramak@molavi.de



THE LAW TECHNOLOGIST

iRights.Lab

Think Tank für die digitale Welt