Maschinelles Lernen verstehen.
Ein Einstieg.

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

Maschinelles Lernen verstehen. Ein Einstieg.

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

Dr. Christian Richter
Referent für Medienbildung (und digitale Kompetenzen)
am Landesinstitut für Schule und Medien Berlin-Brandenburg (LISUM)
Medienwissenschaftler

In Zusammenarbeit mit Cornelia Brückner (LISUM)

Erste Annäherung

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

Beispiele für KI im Alltag

- Empfehlungen bei Amazon und Spotify und Netflix, Chatbots,
- Suchanfragen bei Google, Navigation
- Gesichts- und Bilderkennung
- Empfehlungen medizinischer Behandlungen
- Kreditvergaben, Auswahl von Bewerber:innen
- Einschätzung Gefahr einer wiederholten Straffälligkeit von Kriminellen

Begriff

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

Maschinelles Lernen

- Maschinen finden ohne explizite Programmierung von Regeln selbstständig Muster in Datensätzen
- treffen auf Basis dieser Analyse dann Vorhersagen und Entscheidungen

Voraussetzung ist die Verarbeitung großer Datenmengen

Begriff

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

„Lernstile“

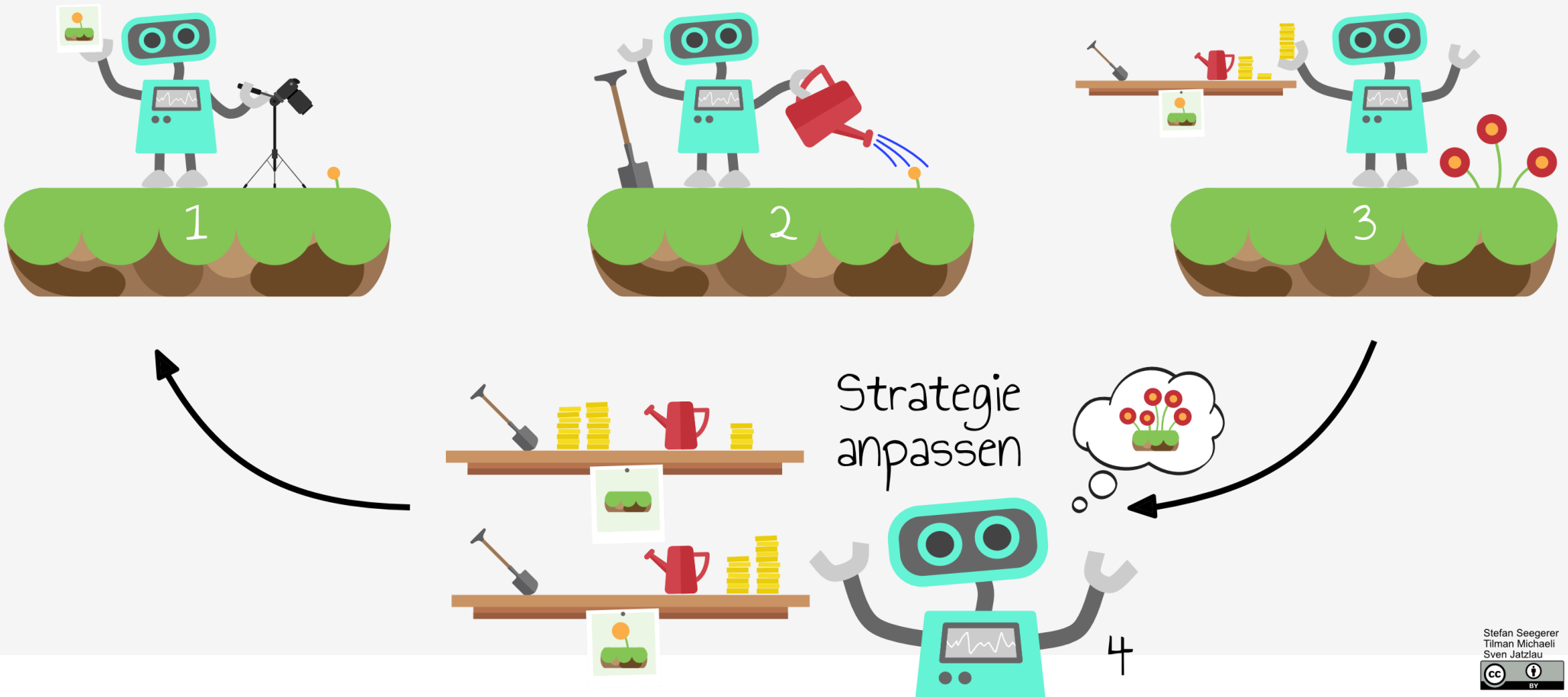
- Reinforced Learning (verstärkendes L)
- Supervised Learning (überwachtes L.)

Reinforced Learning

Zustand erfassen

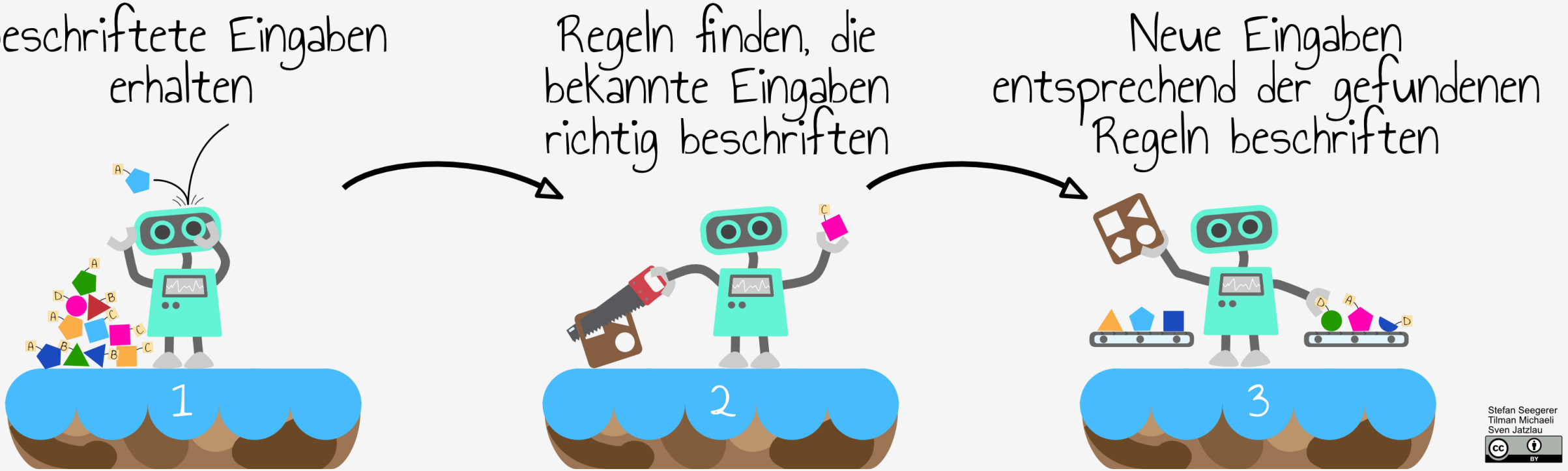
Aktion wählen und durchführen

Belohnung oder Bestrafung erhalten



Supervised Learning

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg



Supervised Learning

- „beschriftete Daten“
- Klassifikation nach Prinzip: „a“ oder „nicht a“
- in der Trainingsphase wird selbständig eine Verbindung zwischen Trainingsdaten und Beschriftung hergestellt
- Ergebnisse werden als Wahrscheinlichkeiten ausgegeben

Datensatz der Passagiere auf der Titanic

LISUM Landesinstitut für Schule und Medien Berlin-Brandenburg

- Mit folgenden Angaben

SURVIVED	Überlebensvariable mit 0=Nein und 1=Ja
PCLASS	Klasse auf dem Schiff 1= Erste, 2=Zweite, 3= Dritte
NAME	Nachname, Title, Vorname(n)
SEX	Geschlecht „male“=männlich, „female“=weiblich
AGE	Alter in Jahren
SIBSP	Anzahl der Geschwister bzw. Partner, die mit an Bord waren
PARCH	Anzahl der Eltern bzw. Kinder, die mit an Bord waren
TICKET	ID Nummer des Tickets
FARE	Preis des Tickets
CABIN	Kabinennummer
EMBARKED	Einstiegshafen C= Cherbourg, Q= Queenstown, S=Southampton

Screenshot von https://www.eoda.de/wissen/blog/oeffentlich-verfuegbare-datensaetze-titanic/

Datensatz der Passagiere auf der Titanic

KI-generiertes Bild

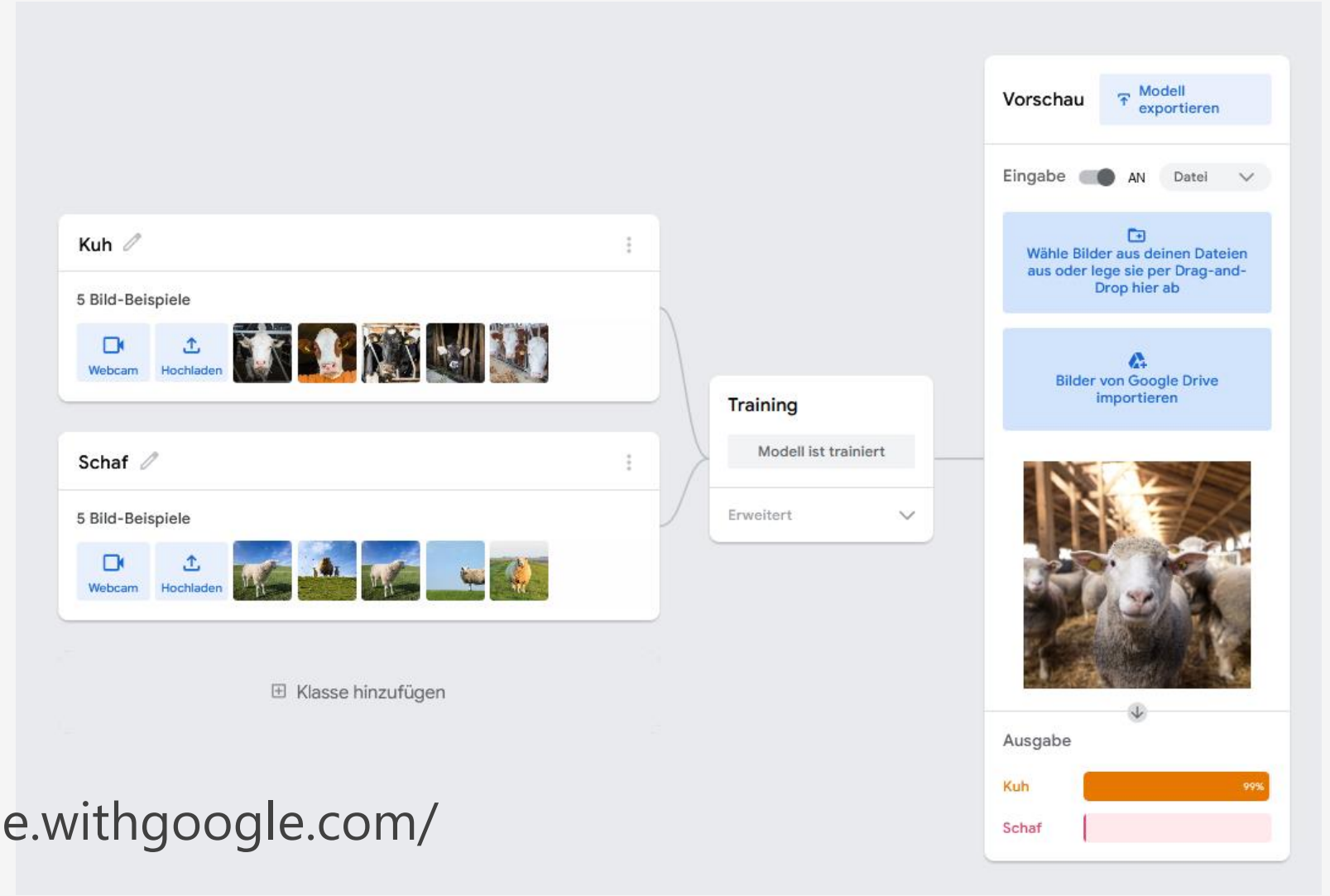
Wie waren die Überlebenschancen von Jack & Rose?

Datensatz der Passagiere auf der Titanic

Wie waren die Überlebenschancen von Jack & Rose?

Modell prognostizierte:
Jack stirbt = 88,9% Wahrscheinlichkeit
Rose überlebt = 96,5% Wahrscheinlichkeit

Teachable Machine



https://teachablemachine.withgoogle.com/

Screenshot

Verzerrung im System

Bias und Fairness in algorithmischen Entscheidungssystemen

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

Unfaire Systeme

- treffen Entscheidungen, bei denen Personen aufgrund ihrer angeborenen oder erworbenen Eigenschaften bevorzugt oder benachteiligt werden
- eine Form von Diskriminierung

Faire Systeme

- stellen eine Gleichbehandlung durch KI-Systeme sicher
- Funktionieren insbesondere ohne Diskriminierung auf Basis von ethnischer Herkunft, Geschlecht, Alter, Religion /Weltanschauung oder sonstiger Indikatoren
- relevant vor allem für KI-Systeme, die Entscheidungen über Personen treffen

Bias

- Verzerrungseffekt in der Datenerhebung oder – verarbeitung
- kann Entscheidungen oder Wahrnehmungen des Systems beeinflussen
- bewusst und unbewusst
- z.B. durch Stereotypen

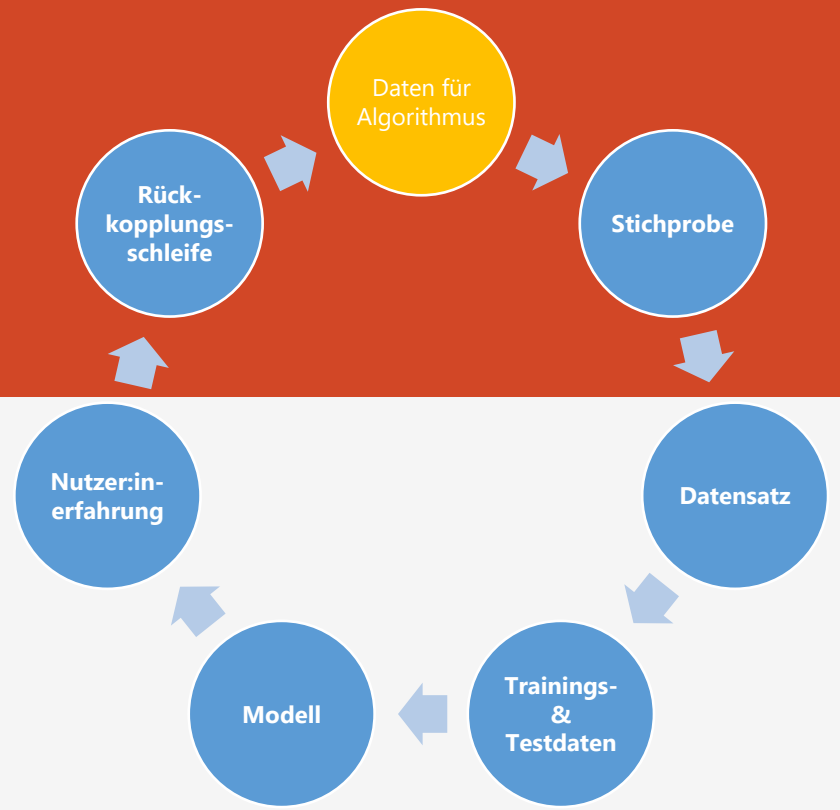
Erste Annäherung

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

- Facebook und LinkedIn setzen KI-Systeme ein, um fragwürdige Inhalte zu filtern. (z.B. gewaltverherrlichende, pornografische oder politisch extreme Bilder, Texte und Videos)
- Welche Schwierigkeiten könnten sich hierbei ergeben?

ML-Prozesses

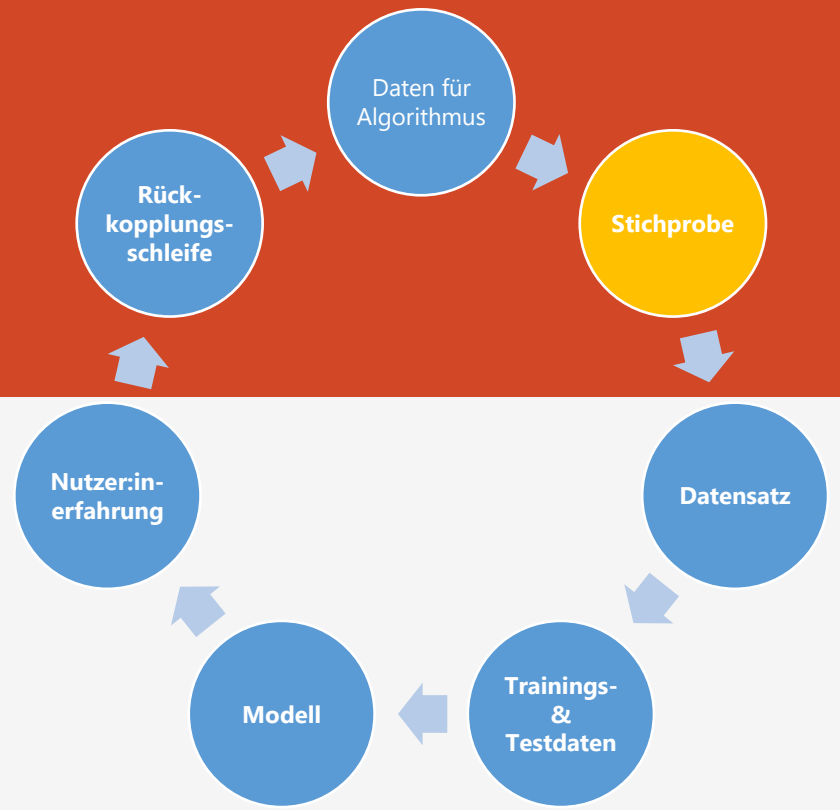
Arten von Bias



Historical Bias

- historische Diskriminierungsmuster
- sind diese Muster im Datensatz enthalten, reproduzieren Systeme diese Muster
- Personen, die in der Vergangenheit Schwierigkeiten hatten, werden durch System weiter benachteiligt
- Bsp.: People of Color in US-Kriminalstatistiken

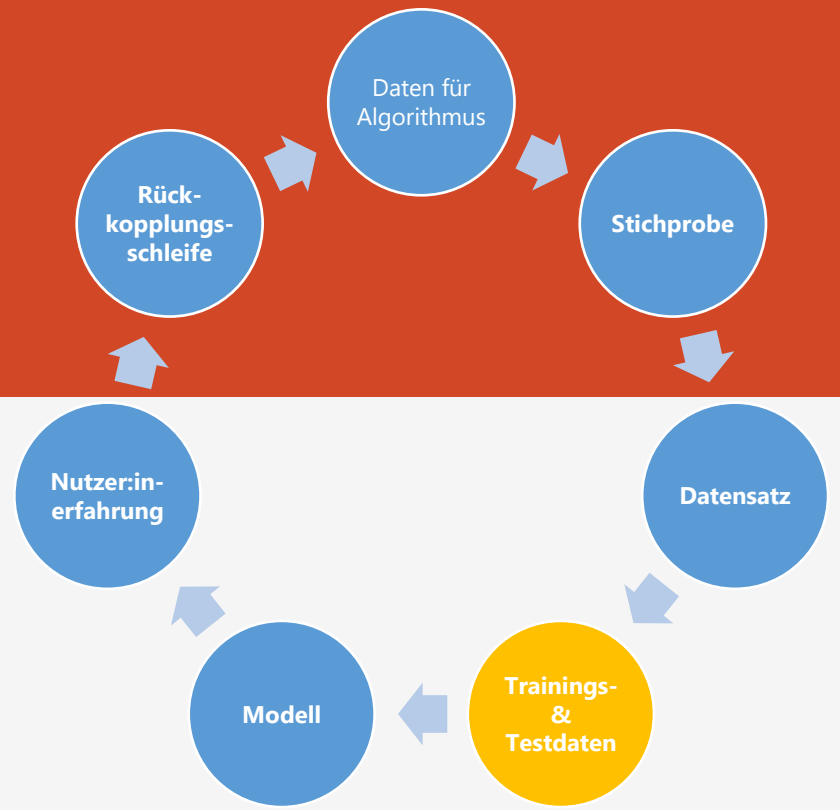
Arten von Bias



Representation Bias

- Nutzung von nicht repräsentativen Stichproben
- Unterrepräsentierte Bereiche verschwinden im System
- Bsp.: beim Training von StableDiffusion & GPT3 wurden nur englischsprachige Quellen genutzt

Arten von Bias



Evaluation Bias

- Testdaten sind nicht repräsentativ
- KI wird unfair trainiert
- Bsp.: Vergleichsdatensatz von Gesichtserkennungssystemen besteht hauptsächlich aus hellhäutigen Menschen

Diskussionsbeispiele

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

- Ein Unternehmen entwickelte ein System, um aus den vielen eingehenden Bewerbungen die geeignetsten automatisiert herauszufiltern. Dafür wurde das System mit den Daten der bereits eingestellten Bewerber:innen trainiert. So sollte es lernen, welche Eigenschaften von der Leitung bei Einstellungen bevorzugt werden. Bei einer Überprüfung kam heraus, dass Frauen durch das System diskriminiert wurden. Wie kam das zustande?

Diskussionsbeispiele

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

- Ein anderes Unternehmen entwickelte ebenfalls ein System, um aus den vielen eingehenden Bewerbungen die geeignetsten automatisiert herauszufiltern. Das System stellt fest, dass Arbeitnehmende, die weiter entfernt vom Firmensitz arbeiten, häufig kündigen. Daher werden Bewerbungen von Personen bevorzugt, die kürze Anfahrtswege haben. Welche Folge hatte diese Entscheidung?

Content-Filter

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

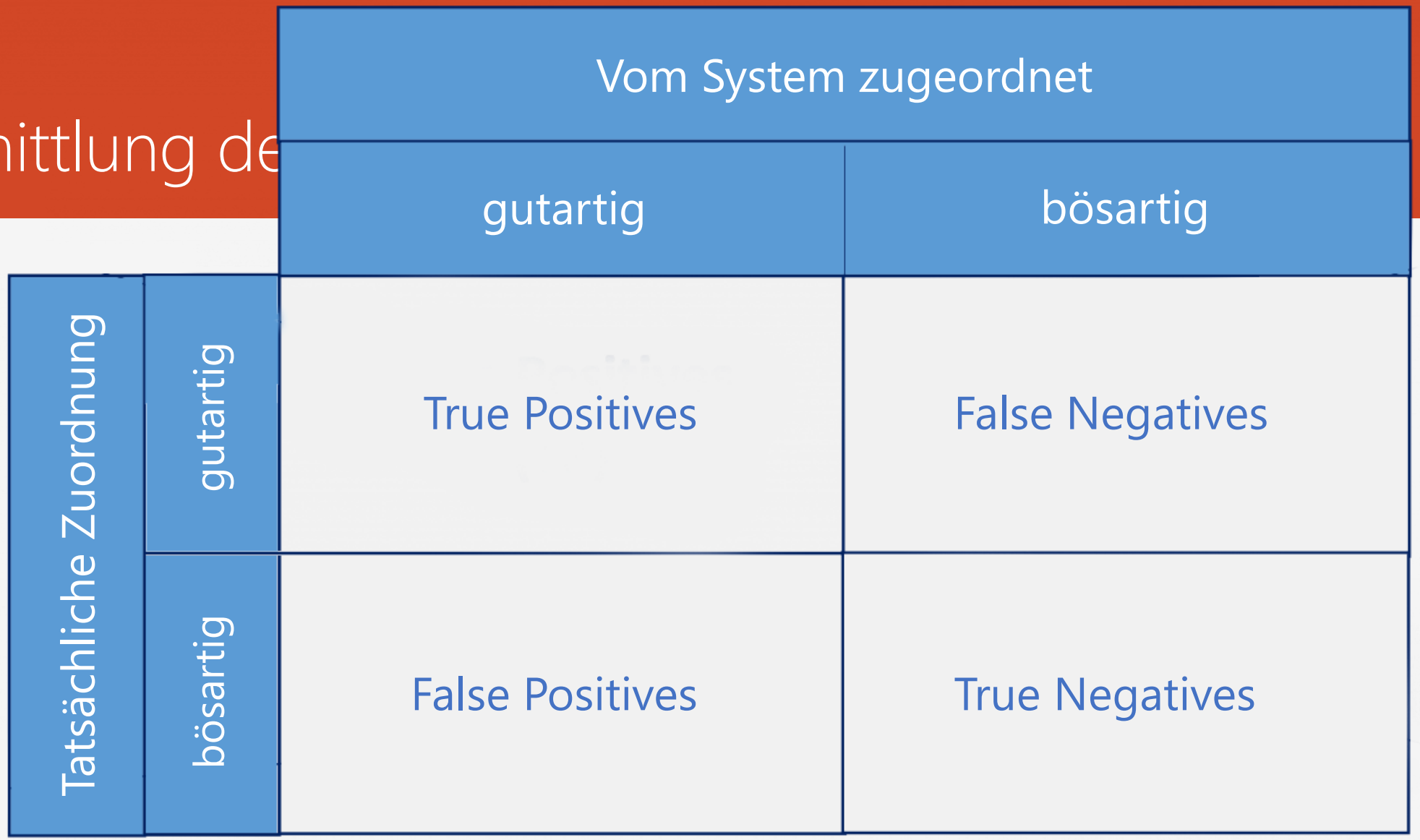
- Facebook und LinkedIn setzen KI-Systeme ein, um fragwürdige Inhalte zu filtern. (z.B. gewaltverherrlichende, pornografische oder politisch extreme Bilder, Texte und Videos)
- Welche Schwierigkeiten könnten sich hierbei ergeben?

Ermittlung der Fairness

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

- Zuordnung der Ergebnisse nach Korrektheit
- also die Zuordnung der Anzahl der richtig und falsch klassifizierten Testdaten
- Testdaten müssen vorher verifiziert sein, Zuordnung muss vorab zweifelsfrei feststehen
- Darstellung in Confusion Matrix

Ermittlung de



Wie kann man Fairness gewährleisten?

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

- Diversität in Programmier-Teams
- Transparenz des KI-Systems, sodass Menschen es nachvollziehen können
- Einsatz von Algorithmen zur Korrektur

Selbstlernkurs

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg

Der Abschnitt „Verzerrung im System“ basiert zu großen Teilen auf dem Online-Angebot:

Bias & Fairness algorithmischer Entscheidungssysteme

des **Instituts für Business Analytics** der Universität Ulm

Prof. Dr. Mathias Klier

https://bias-and-fairness-in-ai-systems.de/

Weitere Quellen

LISUM Landesinstitut für Schule und Medien Berlin-Brandenburg

- Beispiele zu Bias in Bewerbungsprozessen:
  https://www.zeit.de/arbeit/2018-10/bewerbungsroboter-kuenstliche-intelligenz-amazon-frauen-diskriminierung/komplettansicht
- Schlechtes Abschneiden von People of Color bei Gesichtserkennungssoftware
  http://gendershades.org/
- COMPAS - algorithmische Systeme, um Gefahr der wiederholten Straffälligkeit von Kriminellen einzuschätzen
  https://netzpolitik.org/2017/radiofeature-algorithmen-als-schicksalsmaschinen/
- Verzerrungen bei Kreditvergabe (Beispiel für Historical Bias)
  https://safe-frankfurt.de/de/aktuelles/safe-finance-blog/details/ki-basiert-ermittelte-kreditausfallrisiken-sind-mit-vorsicht-zu-geniessen.html
- Titanic-Datensatz - Überleben Rose und Jack?
  https://www.eoda.de/wissen/blog/oeffentlich-verfuegbare-datensaetze-titanic/

Materialien

verwendetes Material

https://computingeducation.de/proj-it2school/

weitere Materialien:

https://www.medien-in-die-schule.de/unterrichtseinheiten/machine-learning-intelligente-maschinen/modul-2-wie-funktioniert-machine-learning/

Deep Learning

https://www.science-on-stage.de/sites/default/files/material/machine-learning-in-der-schule_3.-auflage.pdf

nach Datensätzen suchen https://datasetsearch.research.google.com/

LISUM
Landesinstitut für Schule und Medien Berlin-Brandenburg