

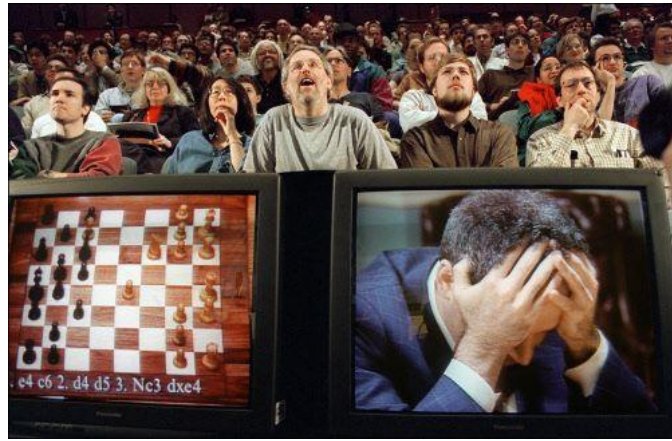
Grenzen künstlicher Intelligenz

...darüber, was KI kann und was nicht

10. Quistorp-Symposium,
22.11.2024
Universität Rostock

Prof. Dr. Gero Szepannek
Statistik, Wirtschaftsmathematik und
Machine Learning
Hochschule Stralsund

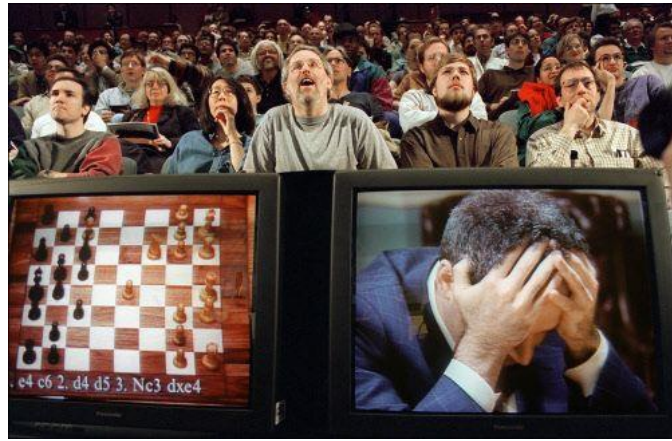




1997
Deep Blue



2017
AlphaGo



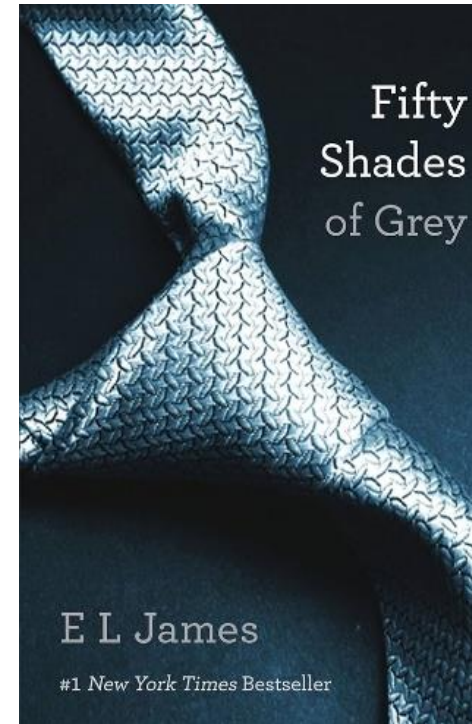
1997
Deep Blue



2017
AlphaGo







„Warum funktioniert KI und warum manchmal nicht?“

**„Integration von KI in unternehmerische Entscheidungen am Bsp.
Kreditvergabe“**

„Können wir KI verstehen?“

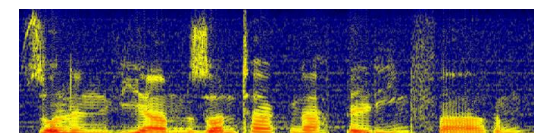
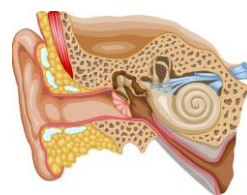
„Gesellschaftliche Herausforderung: Ist KI eigentlich fair?“

Prof. Dr. Gero Szepannek
Statistik, Wirtschaftsmathematik
& Maschinelles Lernen
Hochschule Stralsund



tu technische universität
dortmund





tu technische universität
dortmund

Fraunhofer
IDMT





tu technische universität
dortmund

Fraunhofer
IDMT

Santander
CONSUMER BANK





tu technische universität
dortmund

Fraunhofer
IDMT

Santander
CONSUMER BANK

2016

HOST
Hochschule Stralsund

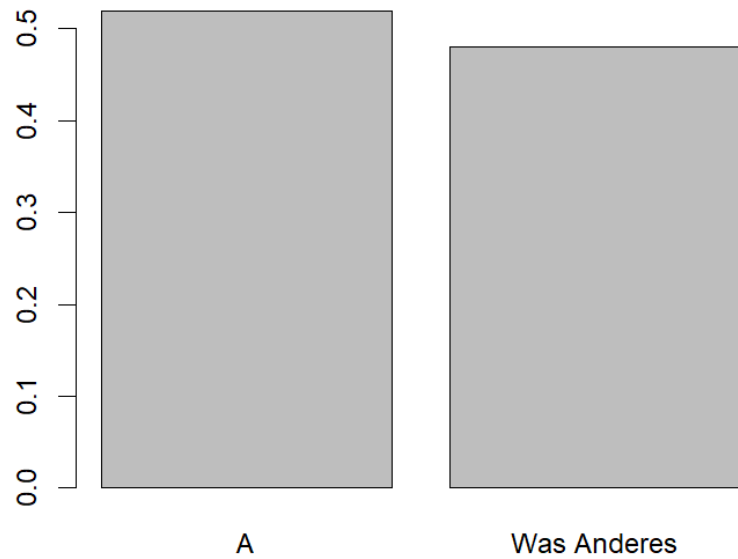
„Warum funktioniert KI und warum manchmal nicht?“

**„Integration von KI in unternehmerische Entscheidungen am Bsp.
Kreditvergabe“**

„Können wir KI verstehen?“

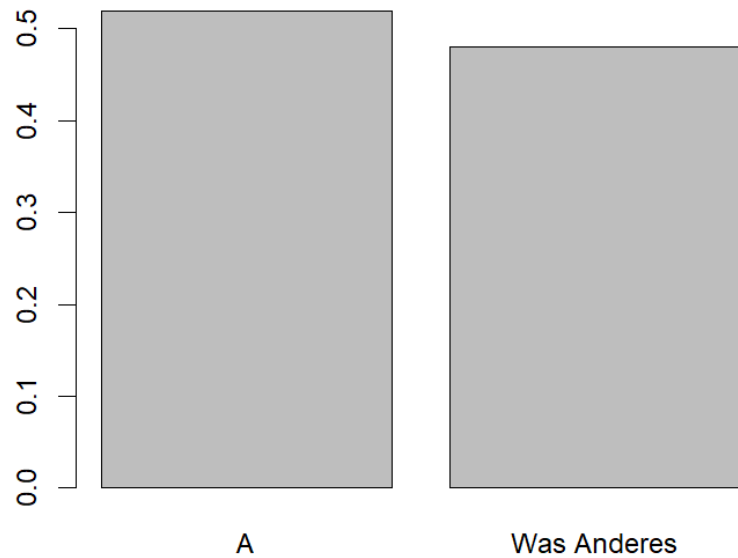
„Gesellschaftliche Herausforderung: Ist KI eigentlich fair?“

Bayes-Regel:



$$\arg \max_A P(A)$$

Bayes-Regel:

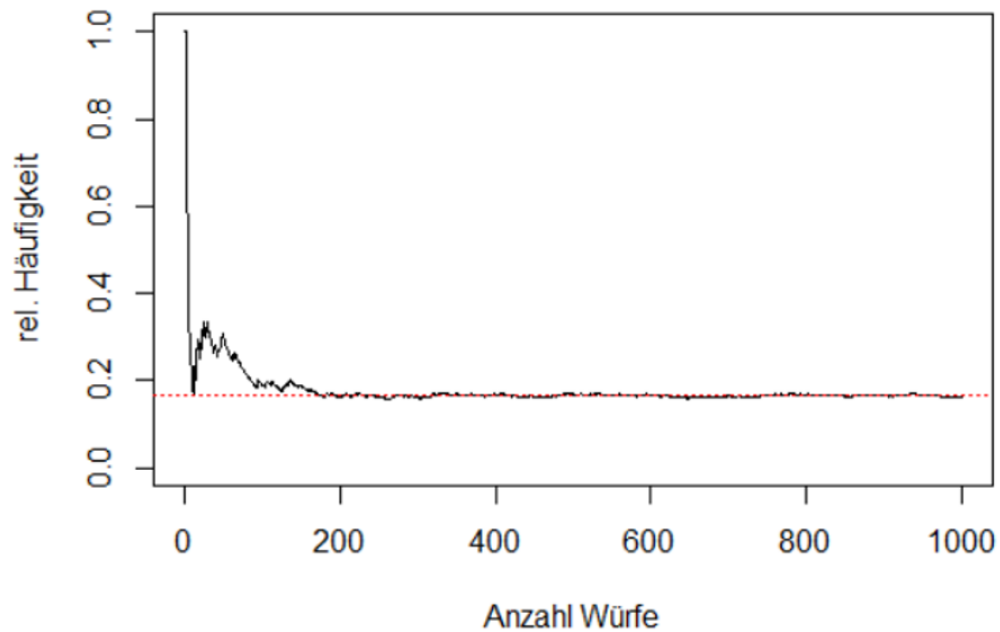


$$\arg \max_A P(A)$$

Gesetz der großen Zahl:

$$P(A) = \lim_{n \rightarrow \infty} f_n(A)$$

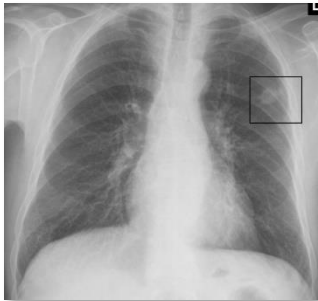
Gesetz der großen Zahl:



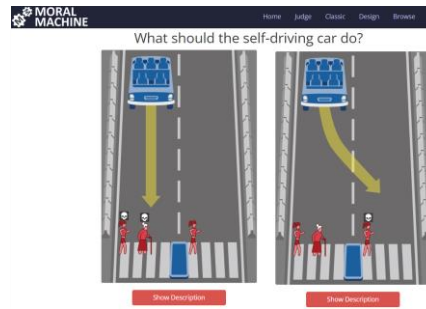
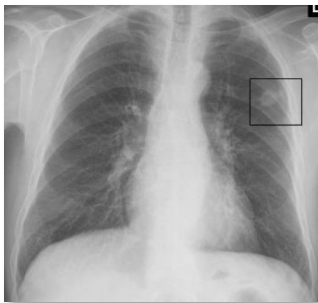
$$P(A) = \lim_{n \rightarrow \infty} f_n(A)$$



***Erfolgreiche KI benötigt zum Lernen
eine ausreichende Menge an Daten!***

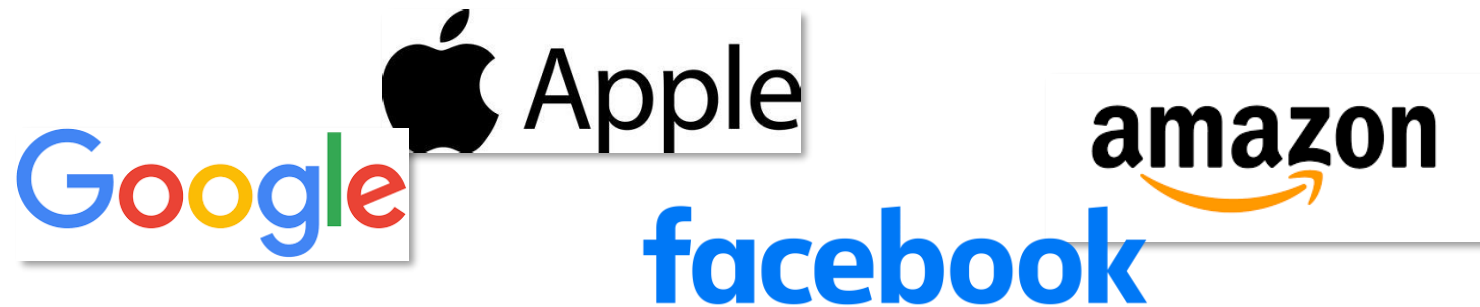
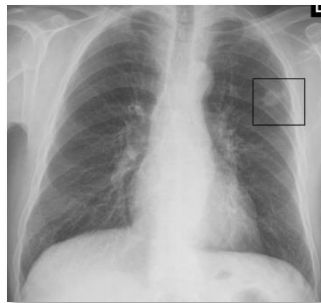


Erfolgreiche KI benötigt zum Lernen eine ausreichende Menge an Daten!



<https://www.moralmachine.net/>

*Erfolgreiche KI benötigt zum Lernen
eine ausreichende Menge an Daten!*

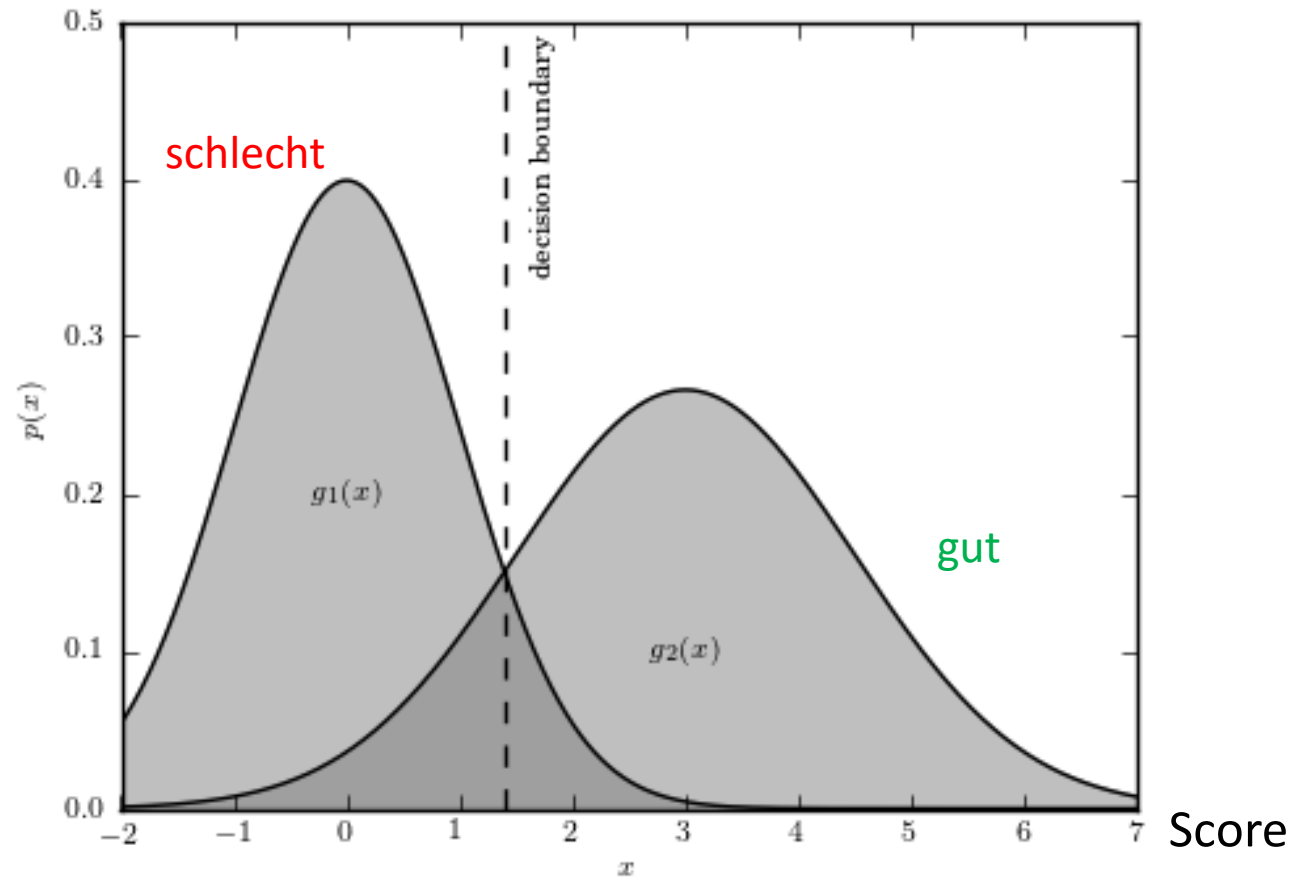


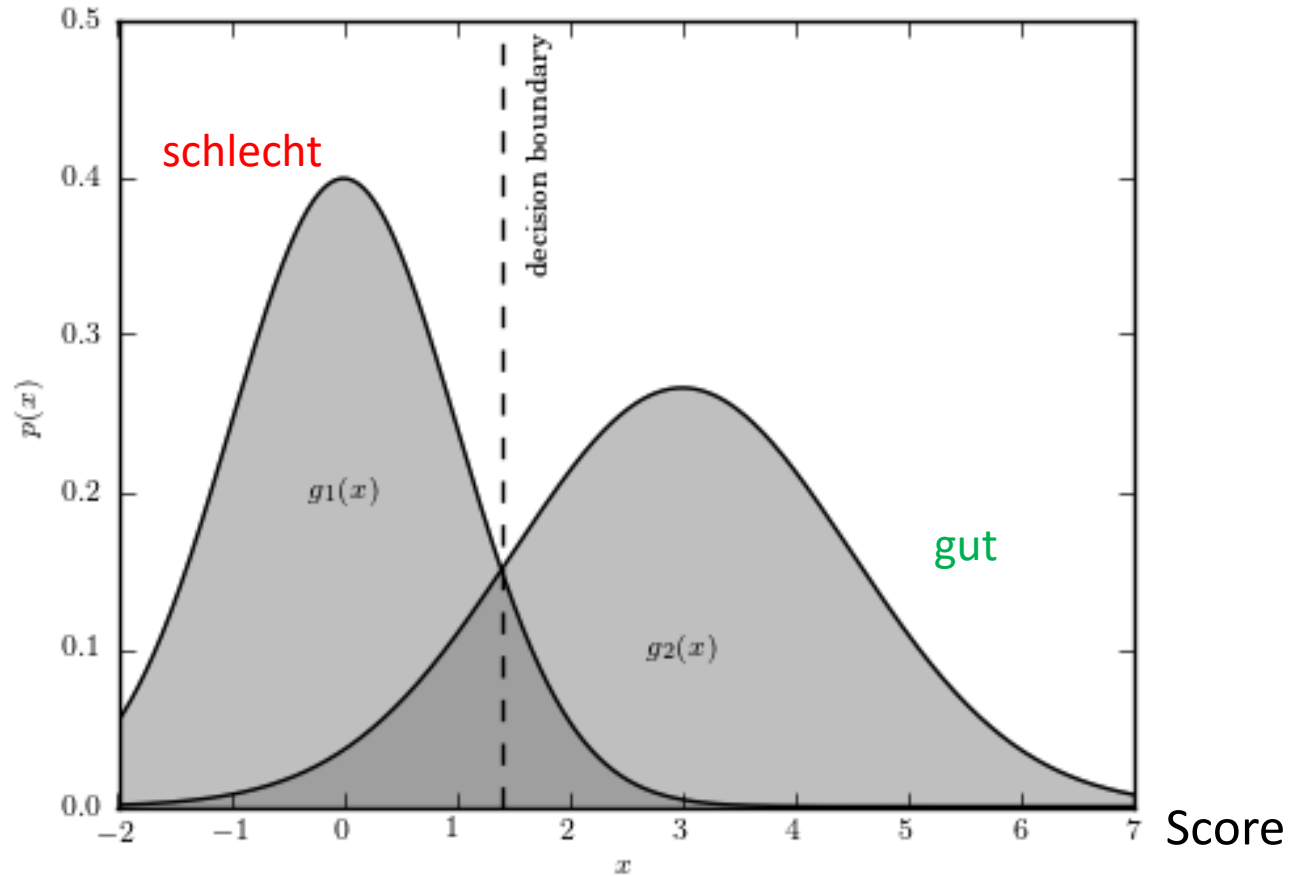
*Erfolgreiche KI benötigt zum Lernen
eine ausreichende Menge an Daten!*



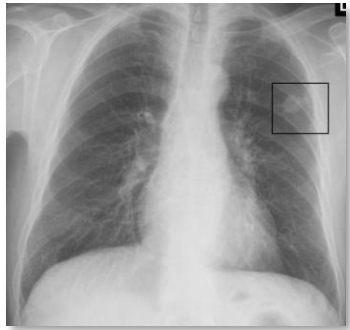


variable	bin	points
basepoints		441
status	no checking account	-17
status	... < 0 DM	-12
status	0 <= ... < 200 DM	11
status	... >= 200 DM / salary for at least 1 year	29
duration	[-Inf,8)	32
duration	[8,18)	10
duration	[18,44)	-5
duration	[44, Inf)	-38
purpose	repairs	-25
purpose	domestic appliances, business, others, radio/television	-16
purpose	retraining	-2
purpose	car (used)	2
purpose	furniture/equipment	17
purpose	car (new), vacation	23
credit history	delay in paying off in the past, critical account/other credits elsewhere	-32
credit history	no credits taken/all credits paid back duly	-3
credit history	existing credits paid back duly till now	0
credit history	all credits at this bank paid back duly	18
savings	unknown/no savings account	-5
savings	... < 100 DM	-3
savings	100 <= ... < 500 DM	13
savings	500 <= ... < 1000 DM, ... >= 1000 DM	15



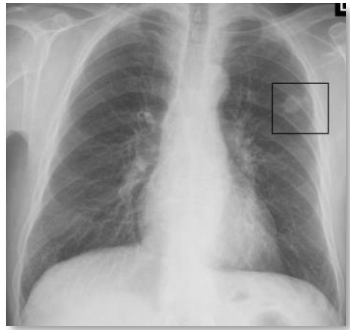


KI macht Fehler!



VS.



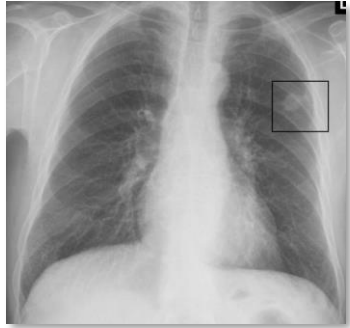


VS.



$$\hat{f}(X) = \arg \min_{f^*(X)} \frac{1}{n} \sum_{i=1}^n L(y_i, f^*(x_i))$$





VS.

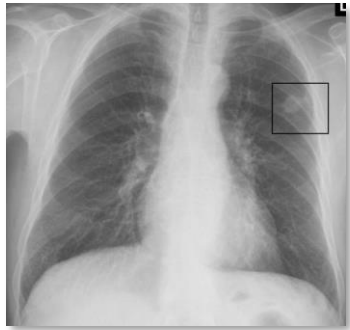


$$\hat{f}(X) = \arg \min_{f^*(X)} \frac{1}{n} \sum_{i=1}^n L(y_i, f^*(x_i))$$

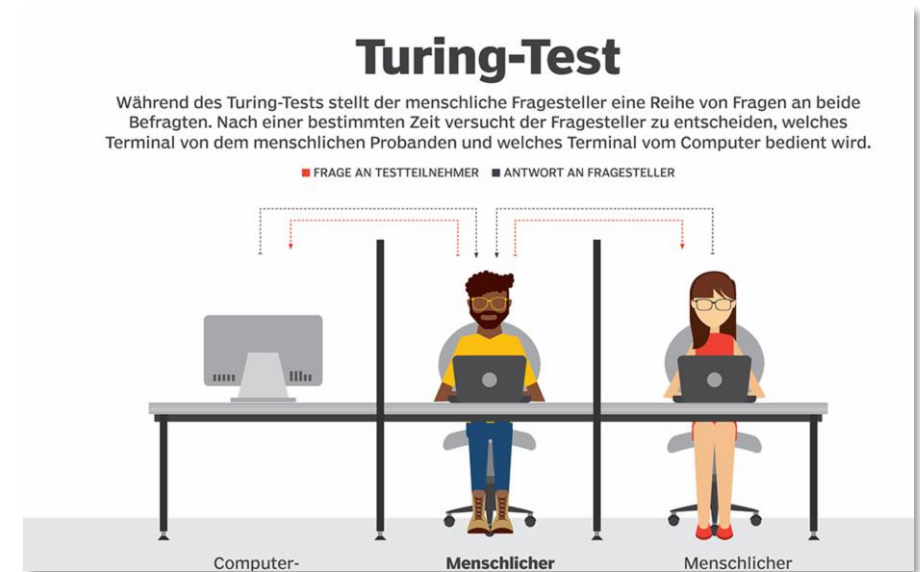


Nein!





VS.

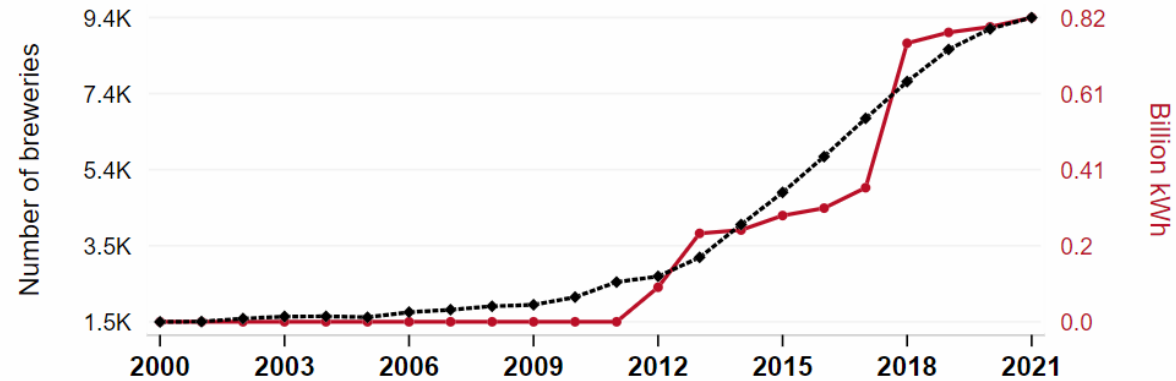


<https://www.computerweekly.com/de/definition/Turing-Test>

The number of Breweries in the United States

correlates with

Solar power generated in Peru

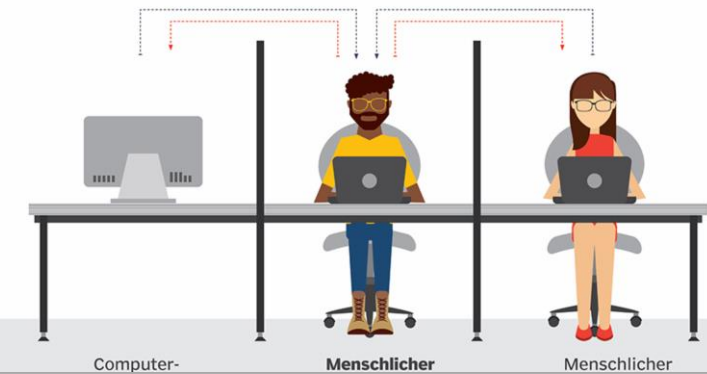


◆--- Number of Breweries in the United States · Source: Brewers Association
 ●— Total solar power generated in Peru in billion kWh · Source: Energy Information Administration
 2000-2021, $r=0.978$, $r^2=0.957$, $p<0.01$ · [tylervigen.com/spurious/correlation/3384](https://www.tylervigen.com/spurious/correlation/3384)

Turing-Test

Während des Turing-Tests stellt der menschliche Fragesteller eine Reihe von Fragen an beide Befragten. Nach einer bestimmten Zeit versucht der Fragesteller zu entscheiden, welches Terminal von dem menschlichen Probanden und welches Terminal vom Computer bedient wird.

■ FRAGE AN TESTTEILNEHMER ■ ANTWORT AN FRAGESTELLER

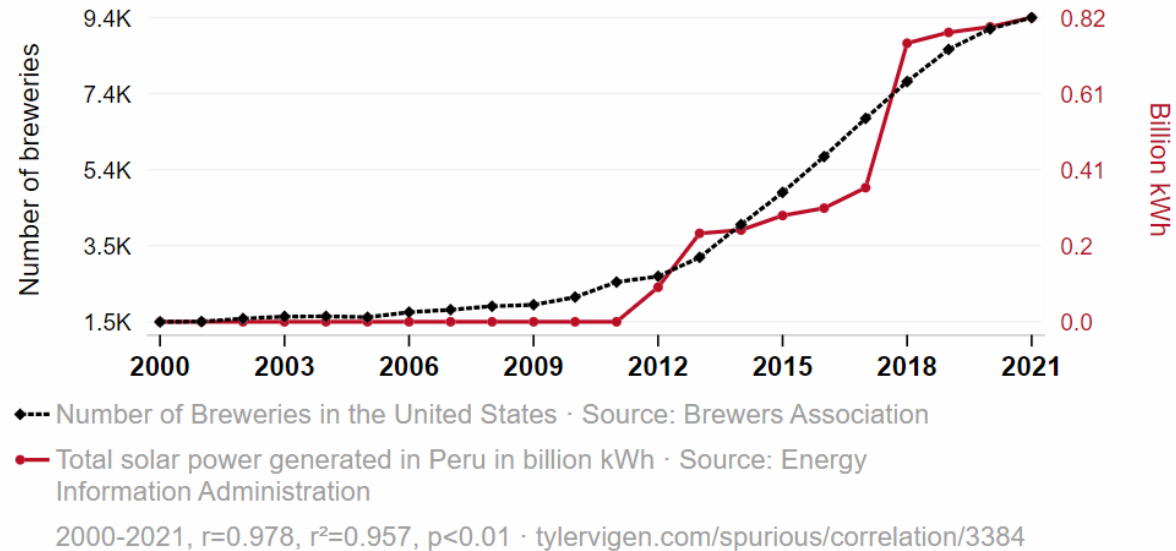


<https://www.tylervigen.com/spurious-correlations>

The number of Breweries in the United States

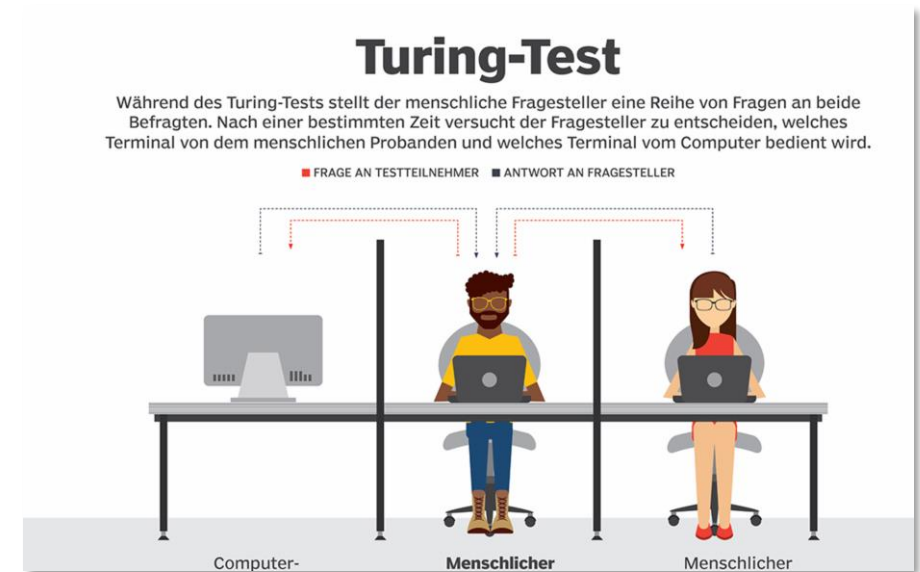
correlates with

Solar power generated in Peru



Show GenAI's made-up explanation

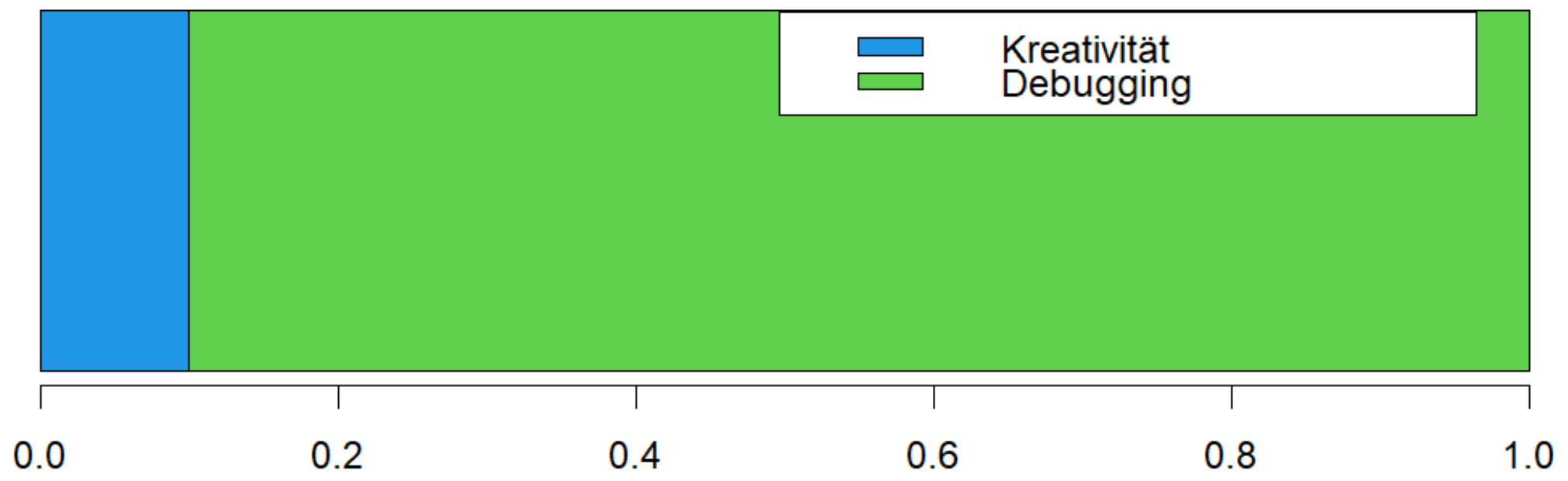
As the number of breweries in the United States increased, so did the demand for beer. This led to a surge in hop production, creating a need for more agricultural labor in Peru. With more people working outdoors, the Peruvian solar power industry experienced a boost in productivity from all the extra sunshine beer as the real power source!



<https://www.tylervigen.com/spurious-correlations>



<https://www.zeit.de/digital/2024-10/programmieren-kuenstliche-intelligenz-technologie/komplettansicht>



<https://www.zeit.de/digital/2024-10/programmieren-kuenstliche-intelligenz-technologie/komplettansicht>

»Mensch und Maschine können gemeinsam Großartiges schaffen«, sagt Meckel. Für Kreativität und echtes »Bio-Denken«, wie sie sagt, sieht sie auch in Zukunft noch viel Bedarf, denn das können die Rechner nicht wirklich. In Kürze werden die Sprachmodelle auf den Computern alles gelesen haben, was es gibt im Internet, und aus diesem Rohmaterial werden sie alle Antworten auf alle Fragen ableiten. Demnach droht also eine gewisse intellektuelle Fadheit: »Wir werden ein permanentes Wiederkäuen von Bestehendem erleben«, sagt Meckel. »Textlich betrachtet beginnt 2026 der globale Inzest.«

(Miriam Meckel, Kommunikationswissenschaftlerin, Universität St. Gallen)

*Regression
to the mean!*



<https://www.spiegel.de/wissenschaft/maschinenintelligenz-schwarze-loecher-wut-auf-suv-die-lese-empfehlungen-der-woche-a-4043e03e-cf32-4b57-bd45-ef34933df816>

„Warum funktioniert KI und warum manchmal nicht?“ ✓

**„Integration von KI in unternehmerische Entscheidungen am Bsp.
Kreditvergabe“ ✓**

„Können wir KI verstehen?“

„Gesellschaftliche Herausforderung: Ist KI eigentlich fair?“

Amazon: KI zur Bewerbungsprüfung benachteiligte Frauen

Eigentlich wollte Amazon eine Software entwickeln, die unter Bewerbern automatisch die besten findet. Der Algorithmus hatte aber unerwünschte Nebenwirkungen.

👍 🔊 🖨️ 💬 430



(Bild: metamorworks/Shutterstock.com)

11.10.2018, 18:32 Uhr Lesezeit: 2 Min.

Von Martin Holland

<https://www.heise.de/news/Amazon-KI-zur-Bewerbungspruefung-benachteiligte-Frauen-4189356.html>

Amazon: KI zur Bewerbungsprüfung benachteiligte Frauen

Eigentlich wollte Amazon eine Software entwickeln, die unter Bewerbern automatisch die besten findet. Der Algorithmus hatte aber unerwünschte Nebenwirkungen.

✓ 🔊 🖨️ 💬 430



(Bild: metamorworks/Shutterstock.com)

11.10.2018, 18:32 Uhr Lesezeit: 2 Min.

Von Martin Holland

Jede KI ist maximal so gut, wie die Daten, mit der man sie füttert!



<https://www.heise.de/news/Amazon-KI-zur-Bewerbungspruefung-benachteiligte-Frauen-4189356.html>

Amazon: KI zur Bewerbungsprüfung benachteiligte Frauen

Eigentlich wollte Amazon eine Software entwickeln, die unter Bewerbern automatisch die besten findet. Der Algorithmus hatte aber unerwünschte Nebenwirkungen.

✓ 🔊 🖨️ 💬 430



(Bild: metamorworks/Shutterstock.com)

11.10.2018, 18:32 Uhr Lesezeit: 2 Min.

Von Martin Holland

Jede KI ist maximal so gut, wie die Daten, mit der man sie füttert!



Garbage in → Garbage out!

<https://www.heise.de/news/Amazon-KI-zur-Bewerbungspruefung-benachteiligte-Frauen-4189356.html>

Amazon: KI zur Bewerbungsprüfung benachteiligte Frauen

Eigentlich wollte Amazon eine Software entwickeln, die unter Bewerbern automatisch die besten findet. Der Algorithmus hatte aber unerwünschte Nebenwirkungen.

✓ 🔊 🖨️ 💬 430

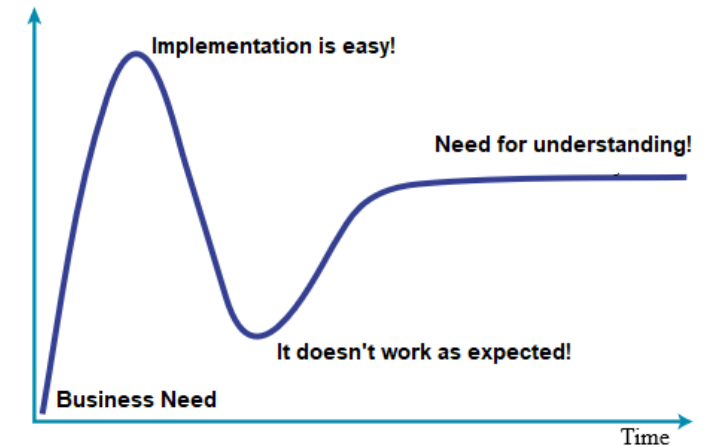


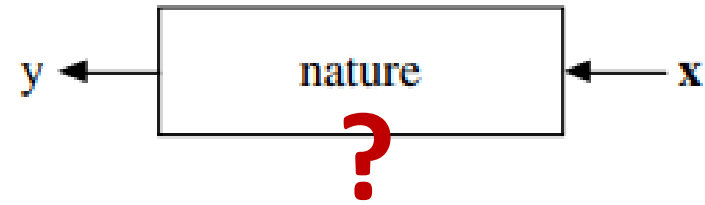
(Bild: metamorworks/Shutterstock.com)

11.10.2018, 18:32 Uhr Lesezeit: 2 Min.

Von Martin Holland

ML Hype Cycle

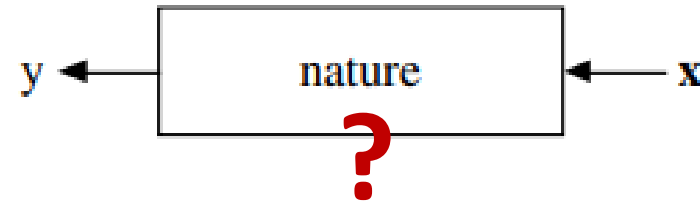




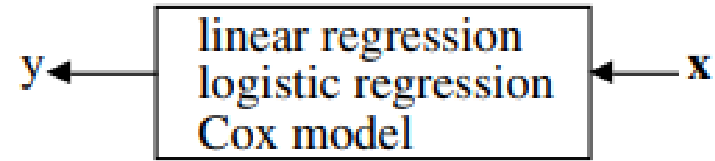
Hypothesen
Daten

Breiman (2001):
Statistical Modelling --
the Two Cultures.

Breiman (2001a):
Statistical Modelling --
the Two Cultures.

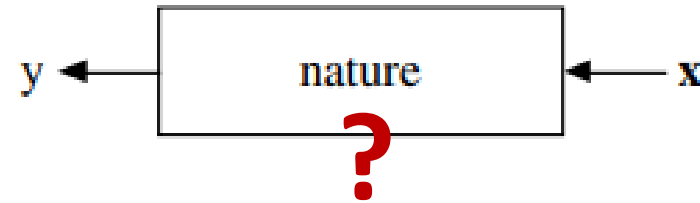


Data Modelling Culture

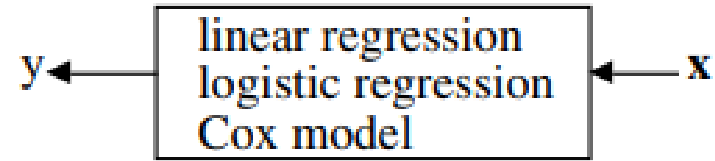


**Deskription und Erklärung von
Kausalzusammenhängen**

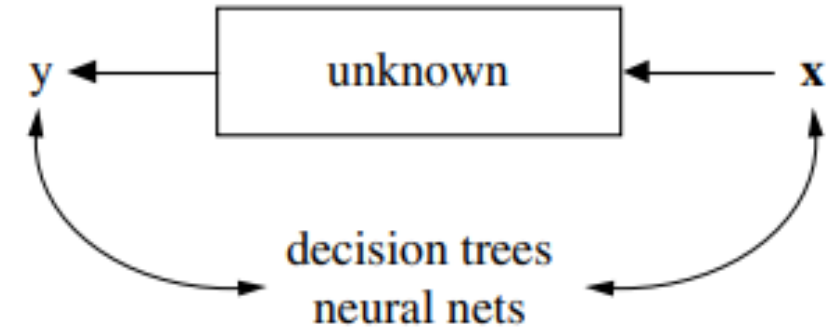
Breiman (2001a):
Statistical Modelling --
the Two Cultures.



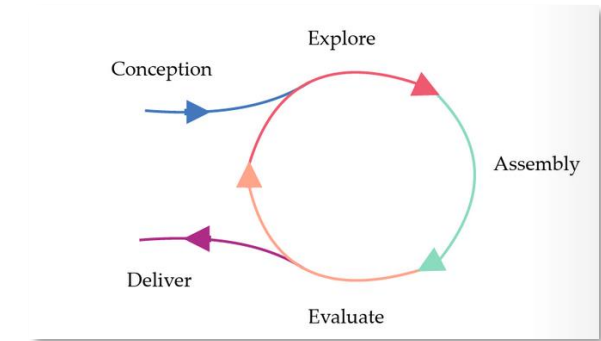
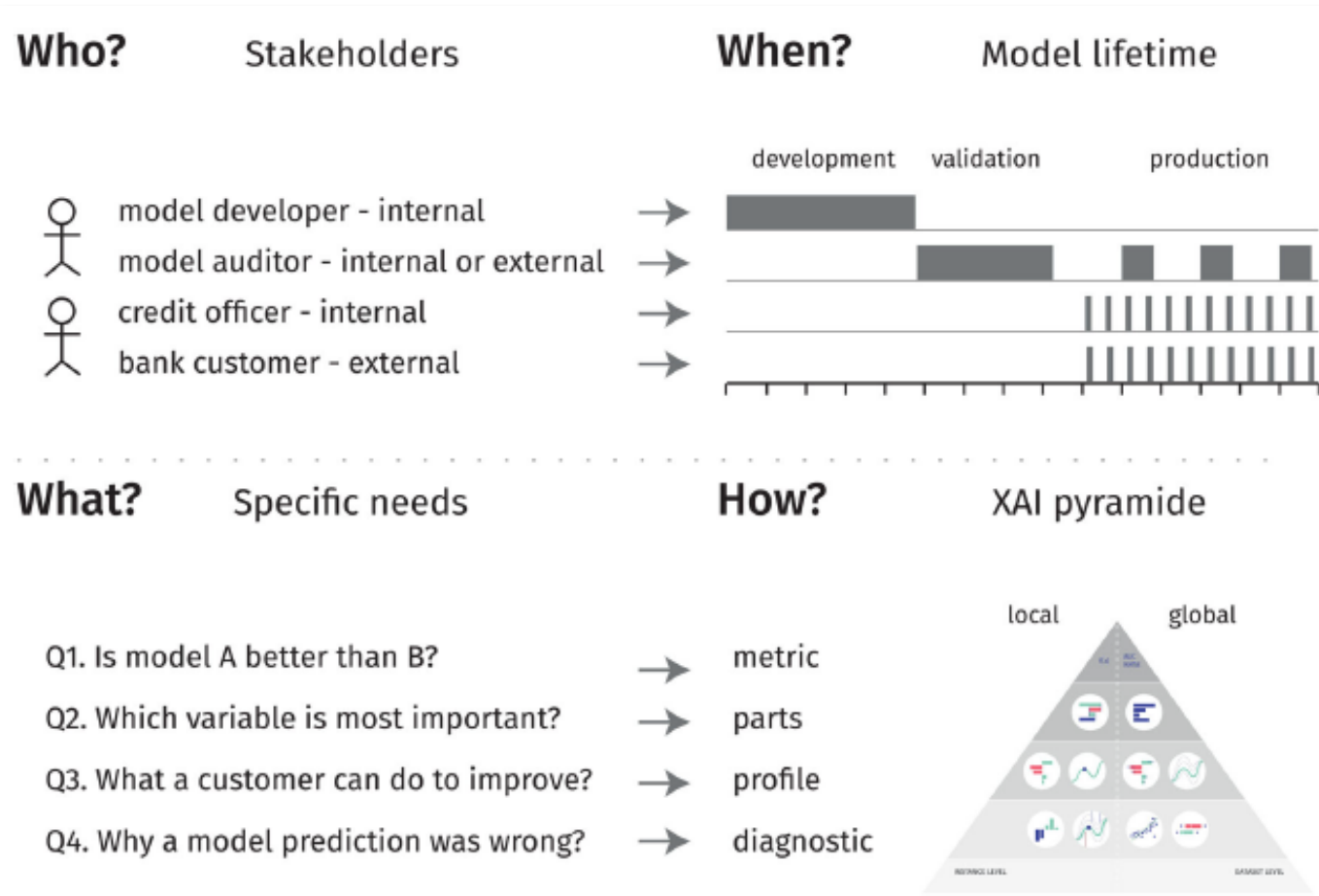
Data Modelling Culture



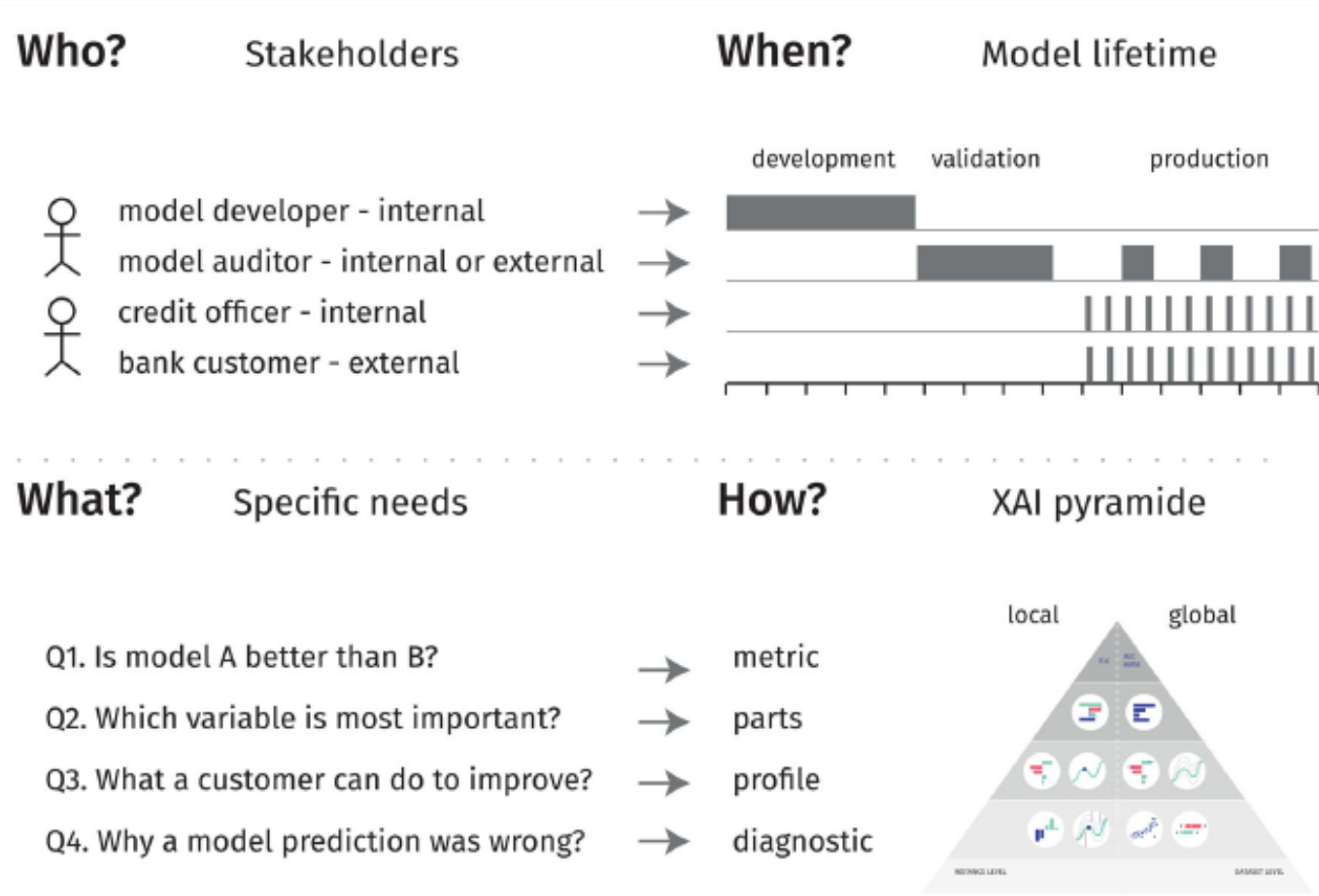
Algorithmic Modelling Culture



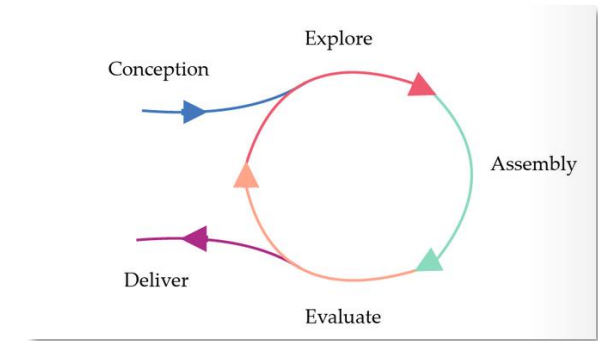
**Deskription und Erklärung von
Kausalzusammenhängen vs. Prädiktion!**



Bücker, Szepannek, Gosiewska und Biecek (2021).



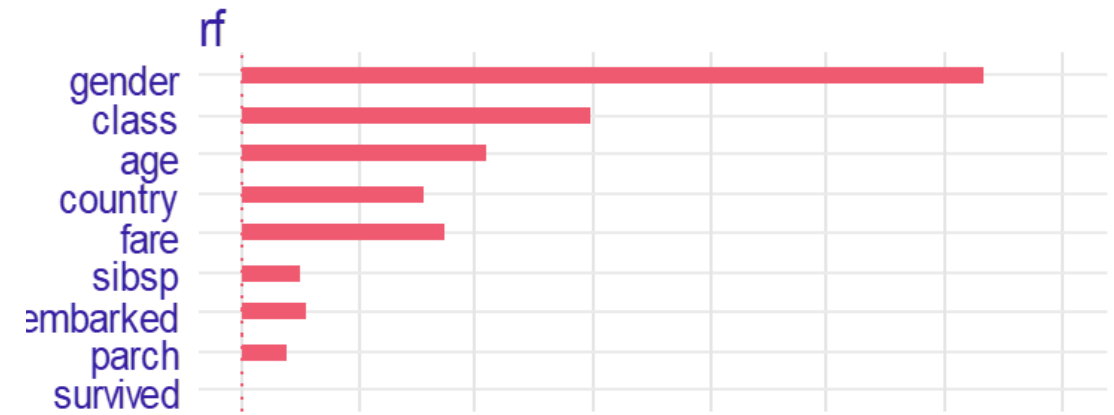
Bücker, Szepannek, Gosiewska und Biecek (2021).

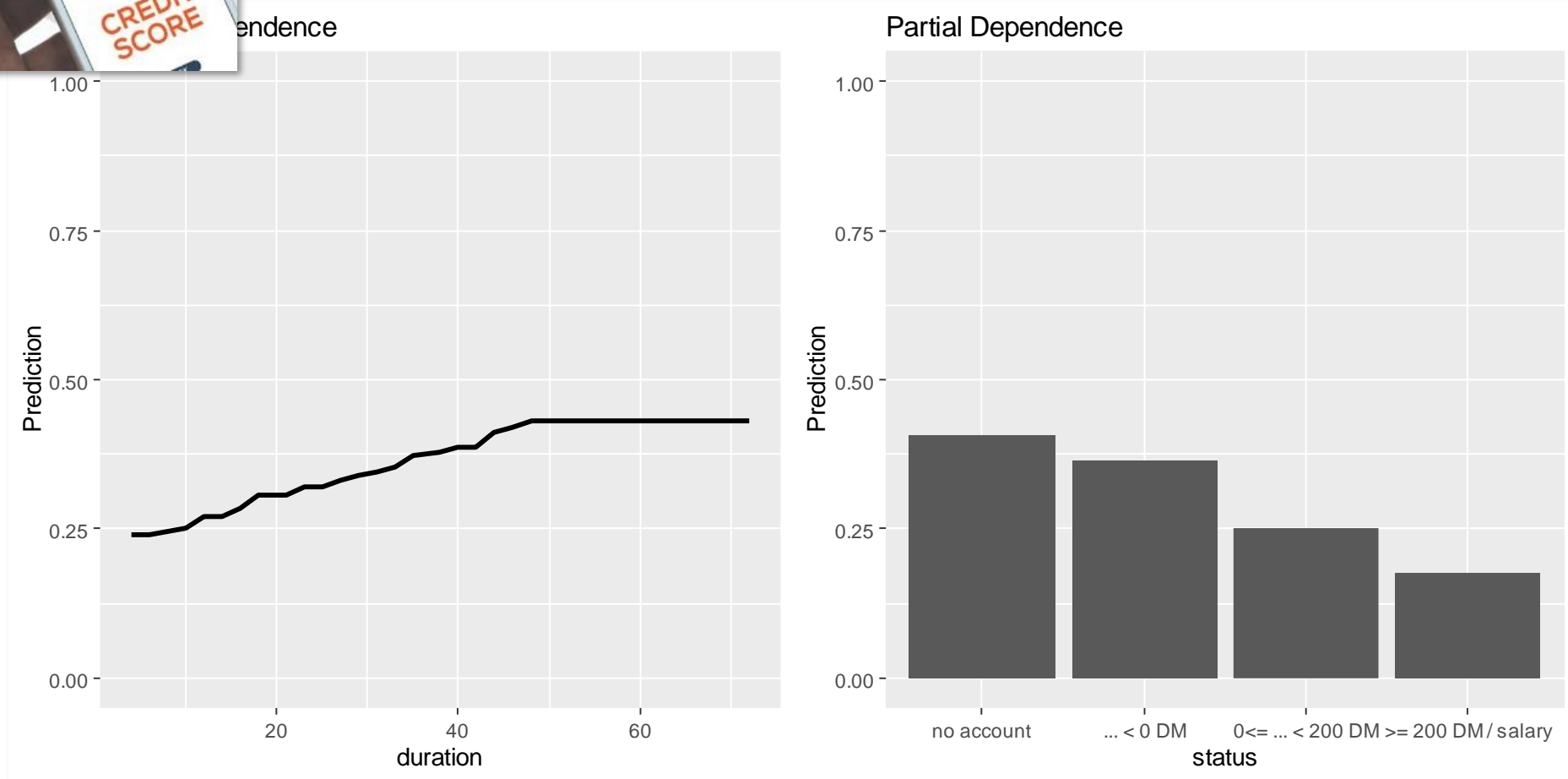


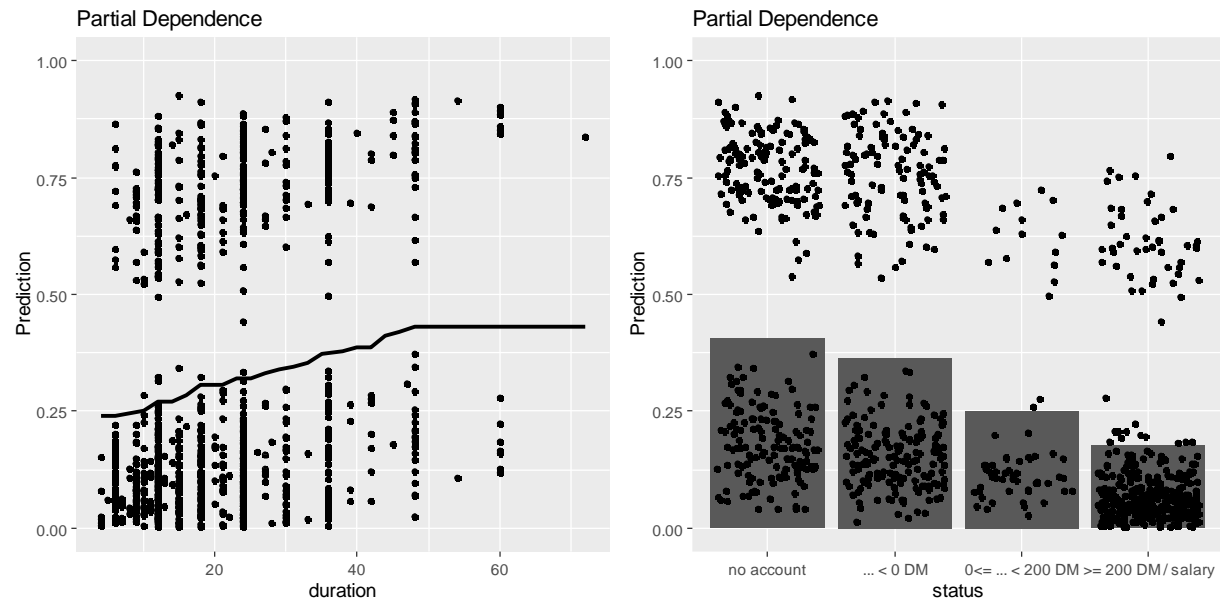
Biecek, Kozak, Zawada und Szepannek (2023)



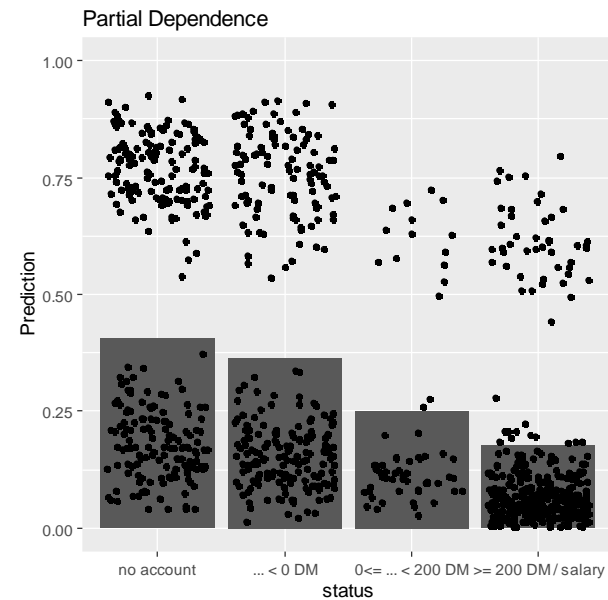
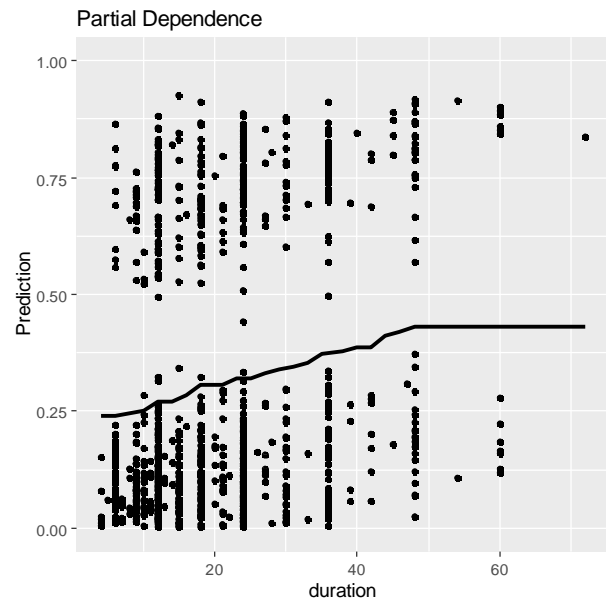
Variable Importance







Szepannek und Lübke (2022).



Variable	Rule (\leq)	Prediction (TRUE)	Prediction(FALSE)
		22.32	
Crime rate	14.14	+0.39	-5.04
<u>Distance to center</u>	1.36	+5.25	-0.15
<u>Longitude</u>	-71.05	+0.76	-0.92
% <u>lower edu</u>	5.23	+3.32	-0.50
% <u>lower edu</u>	14.36	+2.49	-4.77
\emptyset <u>room number</u>	6.81	+0.77	-3.48
\emptyset <u>room number</u>	6.83	-2.76	+12.36
\emptyset <u>room number</u>	7.44	-0.27	+4.64



Szepannek und Lübke (2022).

Szepannek und von Holt (2024).

Amazon: KI zur Bewerbungsprüfung benachteiligte Frauen

Eigentlich wollte Amazon eine Software entwickeln, die unter Bewerbern automatisch die besten findet. Der Algorithmus hatte aber unerwünschte Nebenwirkungen.

👍 🔊 🖨️ 💬 430



(Bild: metamorworks/Shutterstock.com)

11.10.2018, 18:32 Uhr Lesezeit: 2 Min.

Von Martin Holland

FAIR



Name: Long Xie
Alter: 45
Beruf: Journalist
Familienstand: verheiratet
Kinder: 2

210 von 800 Punkten

Mögliche Auswirkungen seines Punktestandes:

- Reiseverbot in Schnellzügen und Flugzeugen.
- Jobverlust.
- Freunde und Angehörige verlieren ebenfalls Punkte, wenn sie sich mit Herrn Long treffen.
- Eingefrorenes Bankkonto.
- Die Nutzung seiner Social-Media-Kanäle ist untersagt.

Bei einem geringen Punktestand im SCS müssen Einwohner mit Sanktionen wie Reiseverboten rechnen. Auch die negativen Konsequenzen können je nach Stadt unterschiedlich ausfallen.

Quelle: <https://www.heise.de/hintergrund/Social-Scoring-in-China-4713878.html>

<https://www.schufa.de/scoring-daten/>

<https://www.huk.de/fahrzeuge/kfz-versicherung/telematik-plus.html>

<https://www.deutschlandfunknova.de/beitrag/social-scoring-wenn-menschen-andere-menschen-bewerten>

(...O'Neil, 2016).



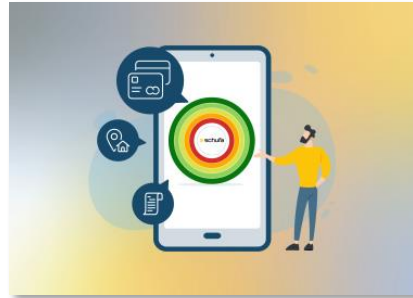
Name: Long Xie
Alter: 45
Beruf: Journalist
Familienstand: verheiratet
Kinder: 2

210 von 800 Punkten

Mögliche Auswirkungen seines Punktestandes:

- Reiseverbot in Schnellzügen und Flugzeugen.
- Jobverlust.
- Freunde und Angehörige verlieren ebenfalls Punkte, wenn sie sich mit Herrn Long treffen.
- Eingefrorenes Bankkonto.
- Die Nutzung seiner Social-Media-Kanäle ist untersagt.

Bei einem geringen Punktestand im SCS müssen Einwohner mit Sanktionen wie Reiseverboten rechnen. Auch die negativen Konsequenzen können je nach Stadt unterschiedlich ausfallen.



Bis zu
30%
sparen

"Ich bin neulich bei meinem Handyprovider auf der Seite gewesen. Und da wurde mir gesagt, ich sei ein Silberkunde. Wie bin ich da rein gekommen? Was habe ich getan?"

Doris Fischer, Professorin für China Economics an der Universität Würzburg

05. Februar 2019

Laut einer neuen Studie fänden es 40 Prozent der Deutschen gut, wenn sie das Verhalten ihrer Mitmenschen positiv oder negativ im Netz bewerten könnten. Warum nur?

Quelle: <https://www.heise.de/hintergrund/Social-Scoring-in-China-4713878.html>

<https://www.schufa.de/scoring-daten/>

<https://www.huk.de/fahrzeuge/kfz-versicherung/telematik-plus.html>

<https://www.deutschlandfunknova.de/beitrag/social-scoring-wenn-menschen-andere-menschen-bewerten>

Unisex-Tarife

Mit geschlechtlichen im Sinne von zweideutigen Angelegenheiten haben Unisex-Tarife nichts zu tun, auch wenn es sich so anhören mag. Bei Unisex-Tarifen geht es nämlich darum, dass das Geschlecht des Kunden bei der Höhe einer Versicherungsprämie keine Rolle spielen darf.

 teilen
  twittern
  senden
 




Unisex-Tarife: Gleiche Prämien für Mann und Frau

 tagesschau
 Sendung verpasst? 
☰


» Wirtschaft » Versicherungen müssen Unisex-Tarife anbieten: Der kleine Unterschied spielt keine Rolle mehr



Versicherungen müssen Unisex-Tarife anbieten

Der kleine Unterschied spielt keine Rolle mehr

Stand: 20.12.2012 20:01 Uhr

In der EU dürfen keine Versicherungen mehr verkauft werden, bei denen Männer und Frauen wegen ihres Geschlechts unterschiedliche Preise zahlen müssen. Unisex-Tarife sind dann Pflicht. Hintergrund ist ein EuGH-Urteil. Es gibt zwar noch kein Gesetz dafür in Deutschland - das macht aber nichts.

Unisex-Tarife

Mit geschlechtlichen im Sinne von zweideutigen Angelegenheiten haben Unisex-Tarife nichts zu tun, auch wenn es sich so anhören mag. Bei Unisex-Tarifen geht es nämlich darum, dass das Geschlecht des Kunden bei der Höhe einer Versicherungsprämie keine Rolle spielen darf.

teilen
 twittern
 senden



Unisex-Tarife: Gleiche Prämien für Mann und Frau

tagesschau
 Sendung verpasst? ▶
☰

Wirtschaft
Versicherungen müssen Unisex-Tarife anbieten: Der kleine Unterschied spielt keine Rolle mehr



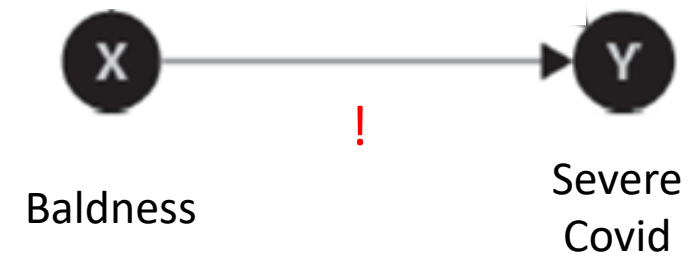
Versicherungen müssen Unisex-Tarife anbieten

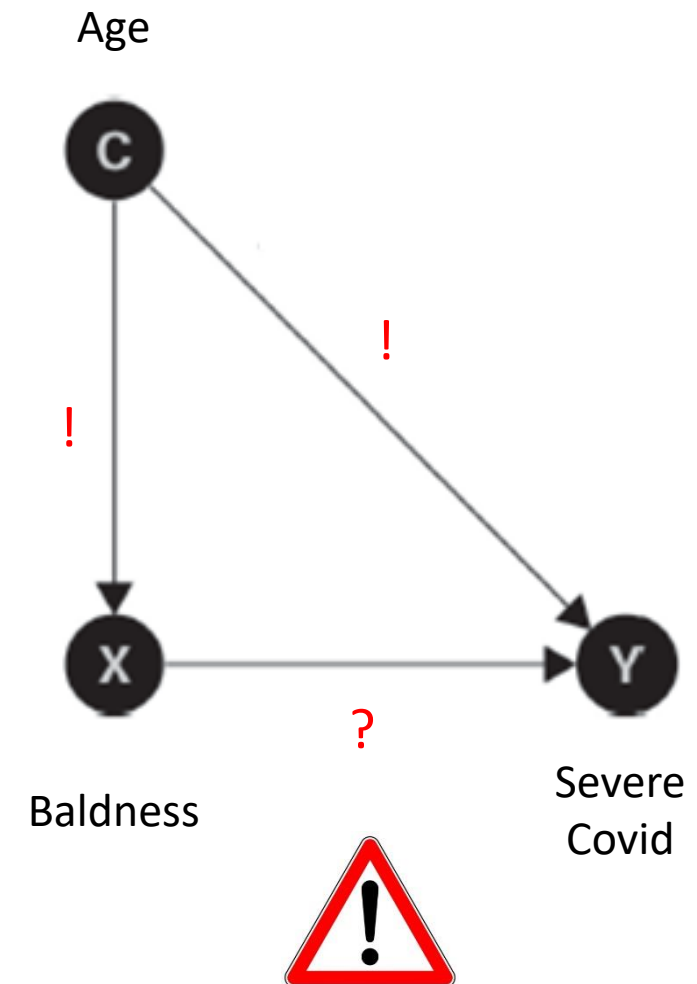
Der kleine Unterschied spielt keine Rolle mehr

Stand: 20.12.2012 20:01 Uhr

In der EU dürfen keine Versicherungen mehr verkauft werden, bei denen Männer und Frauen wegen ihres Geschlechts unterschiedliche Preise zahlen müssen. Unisex-Tarife sind dann Pflicht. Hintergrund ist ein EuGH-Urteil. Es gibt zwar noch kein Gesetz dafür in Deutschland - das macht aber nichts.

„Fairness by unawareness“

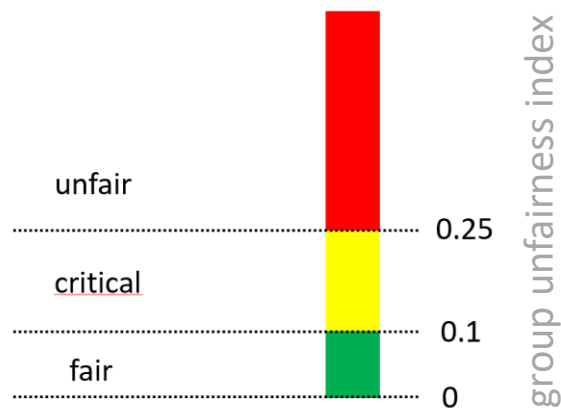




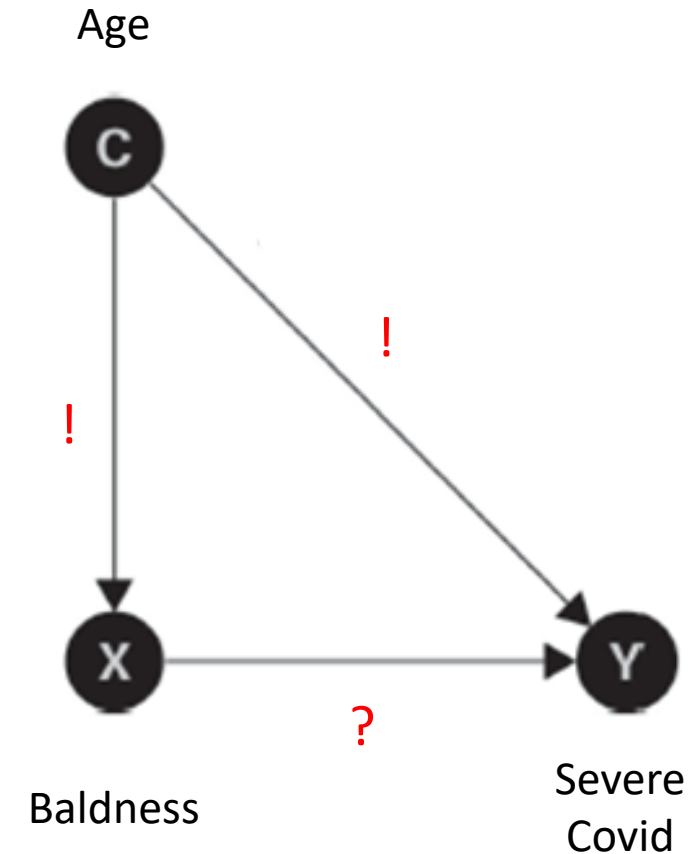
➔ Remove effect of age on baldness!

FairLearning algorithm (Kusner et al., 2017):

1. Regress X on P .
2. Calculate residuals $E = X - \hat{X}$.
3. Model Y by E .



Szepannek und Lübke (2021)





traditional

bin	dist	%bad	woe	points
no checking account	0.281	0.452	0.723	-17
... < 0 DM	0.263	0.397	0.497	-12
0 ≤ ... < 200 DM	0.063	0.205	-0.442	11
... ≥ 200 DM / salary	0.393	0.105	-1.222	29



FAIR

traditional

fairness corrected model

bin	dist	%bad	woe	points	dist f	%bad f	woe f	points f	dist m	%bad m	woe m	points m
no checking account	0.281	0.452	0.723	-17	0.142	0.467	1.343	-34	0.397	0.447	0.527	-13
... < 0 DM	0.263	0.397	0.497	-12	0.145	0.522	1.117	-28	0.360	0.355	0.301	-8
0 ≤ ... < 200 DM	0.063	0.205	-0.442	11	0.098	0.194	0.178	-4	0.034	0.231	-0.638	16
... ≥ 200 DM / salary	0.393	0.105	-1.222	29	0.615	0.103	-0.602	15	0.209	0.112	-1.418	36



FAIR

traditional

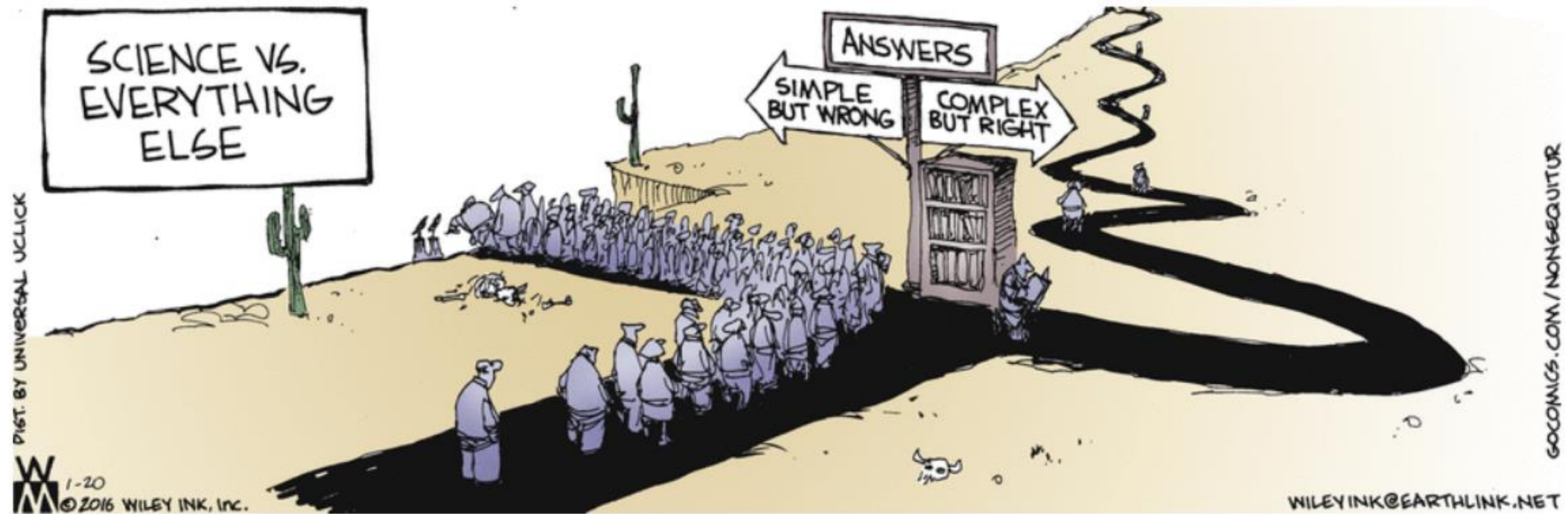
fairness corrected model

bin	dist	%bad	woe	points	dist f	%bad f	woe f	points f	dist m	%bad m	woe m	points m
no checking account	0.281	0.452	0.723	-17	0.142	0.467	1.343	-34	0.397	0.447	0.527	-13
... < 0 DM	0.263	0.397	0.497	-12	0.145	0.522	1.117	-28	0.360	0.355	0.301	-8
0 ≤ ... < 200 DM	0.063	0.205	-0.442	11	0.098	0.194	0.178	-4	0.034	0.231	-0.638	16
... ≥ 200 DM / salary	0.393	0.105	-1.222	29	0.615	0.103	-0.602	15	0.209	0.112	-1.418	36

...fairness correction means:

1. **Same income** (i.e. ceterus paribus) but **different points**!
2. In order to correct for it, **gender has to be assessed** ...What about data protection?





„Warum funktioniert KI und warum manchmal nicht?“ ✓

**„Integration von KI in unternehmerische Entscheidungen am Bsp.
Kreditvergabe“ ✓**

„Können wir KI verstehen?“ ✓

„Gesellschaftliche Herausforderung: Ist KI eigentlich fair?“ ✓

- I. KI ist weder schwarz noch weiß!
- II. KI lernt, spezifische Aufgaben zu lösen!
- III. KI macht Fehler!
- IV. Erfolgreiche KI benötigt viele Daten!
- V. Garbage in \Rightarrow Garbage out!
- VI. Die zwei Kulturen der Modellierung!
- VII. Dringender Bedarf, KI zu verstehen!
- VIII. ...aber: XAI ist nicht die Lösung aller Probleme!
- IX. Faire KI ist grundsätzlich möglich, aber...



- **Biecek, P., Kozak, A., Zawada, A., Szepannek, G. (2023):** Per Anhalter durch die Galaxis des verantwortungsvollen maschinellen Lernens, SmarterPoland, ISBN: 978-8365291189.
- **Breiman, L. (2001a):** Statistical Modeling: The Two Cultures. *Statist. Sci.* 16 (3). 199 - 231, DOI: 10.1214/ss/1009213726
- **Breiman, L. (2001b):** Random Forests, *Machine Learning* 45(1), 5 – 32.
- **Bücker, M., Szepannek, G., Gosiewska, A. Biecek, P. (2021):** Transparency, Auditability and eXplainability of Machine Learning Models in Credit Scoring, *Journal of the Operation Research Society*.
- **Friedman, J. (2001):** Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29:1189–1232.
- **Lübke, K.; Gehrke, M.; Horst, J., Szepannek, G. (2020):** Why We Should Teach Causal Inference: Examples in Linear Regression with Simulated Data, *Journal of Statistics Education*, 28 (2), 133-139, DOI:
- **Kusner, M., Loftus, J., Russell, C., and Silva, R. (2017).** Counterfactual fairness. in *Proc. 31st int. Conf. Neural Information Processing Systems NIPS'17*. (Red Hook, NY, USA: Curran Associates Inc.), 4069–4079.
- **O'Neil, C. (2016):** *The Weapons of Math Destruction*, Brown books.
- **Szepannek, G., von Holt, B.-H. (2024):** Can't See the Forest for the Trees – Analyzing Groves for Random Forest Explanation, *Behaviormetrika*, DOI: 10.1007/s41237-023-00205-2.
- **Szepannek, G., Luebke, K. (2021).** Facing the Challenges of Developing Fair Risk Scoring Models, *Frontiers in Artificial Intelligence* 4, DOI: 10.3389/frai.2021.681915.
- **Szepannek, G., Luebke, K. (2022):** Explaining Artificial Intelligence with Care -- Analyzing the Explainability of Black Box Multiclass Machine Learning Models in Forensics, *Künstliche Intelligenz*, DOI : 10.1007/s13218-022-00764-8.

*Vielen Dank für Ihre
Aufmerksamkeit!*