

Künstliche Intelligenz & Big Data: Freiheit ohne Ende oder Ende der Freiheit?

Ansgar Steland

Institut für Statistik u. Wirtschaftsmathematik und AI Center, RWTH Aachen,

Stand: Dezember 2022

Gliederung

- I. Was kann KI/ML?
- II. Fallstricke...
- III. KI-Förderung, Auswirkungen auf die Wirtschaft
- IV. Gesellschaftliche Folgen
- V. Regulierung: EU AI Act
- VI. Kritik

Geschichte

- Geschichte der KI ist Geschichte der Mathematik
- 9. Jh: „Al-Jabr“ (-> Algebra) Buch v. Al-Kwarizmi (-> Algorithmus)
- Webstuhl benutzte Lochkarten, Übernahme durch erste Computer
- Charles Babbage & Ada Lovelace: Erste mech. Rechenmaschine „Analytical Engine“ (1833)
- Modell für Neuron, McCulloch & Pitts, (1943)
- 1940er/50er: Alan Turing, Turing-Test (1950, Nicht-Unterscheidbarkeit im Dialog)
- 1956: Begriff „Künstliche Intelligenz“ John McCarthy
- 1962: Visuelle Verarbeitung bei Katzen (Nobelpreis, Hubel & Wiesel) -> Modell ‚Deep Learning‘
- 1966: Eliza von Joseph Weizenbaum (erste Chatbot)
- 1997: Deep Blue von IBM gewinnt im Schach gegen Weltmeister Garri Kasparov
- 2001: Random Forest (Breiman)
- 2006: Extreme Learning Machines
- 2007: Boosting, Bagging (Freund, Schapire, Breiman)
- 2011: Watson von IBM gewinnt im Quiz Jeopardy!
- 2016: AlphaGo von Google gewinnt in Go (Training aus historischen Spielen und Spielen gegen sich selbst)
- 2020: AlphaFold: Vorhersage von Protein-Strukturen (Proteine: Kette von Aminosäuren, bildet 3D-Struktur, die relevant für biologische Struktur ist.
- 2022: ChatGPT, Galactica

I. Was kann AI?

<https://cdn.openai.com/outpainting/draft-20220829b/girl-with-a-pearl-earring-bts-cropped.mp4>

Hoffnungen (und Versprechungen...)

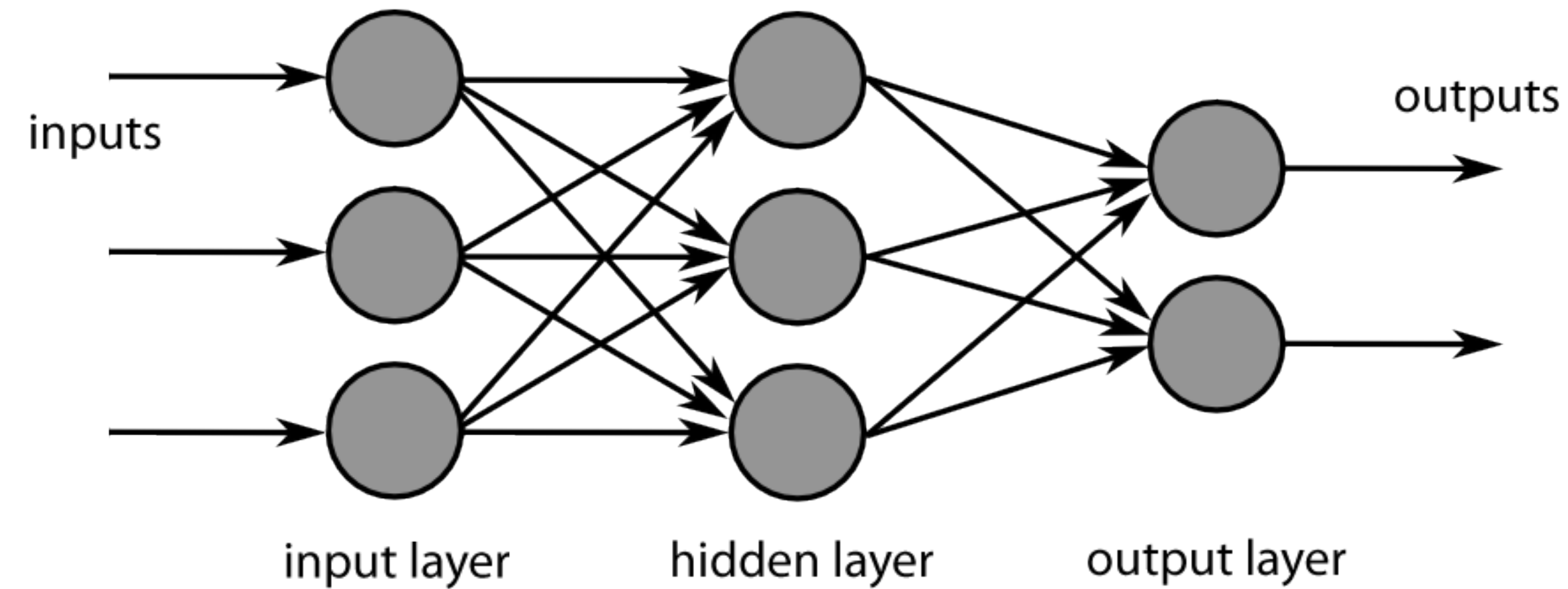
- Klima besser verstehen und prognostizieren.
- Neue und bessere Medikamente und Therapien, Chirurgie-Roboter.
- Optimierung von Energienetzen, Verkehr.
- KI-Assistenten für Beschäftigte
- KI als Wissenschaftler/in: Lösung offener Probleme, KI-generierte Forschung.
- Pflege: autonomer Rollator, Robotikarm, Exoskelett, ...
- Lernassistenten für individualisiertes Lernen.
- Das Wissen der Menschheit und Experten-Intelligenz allen Menschen verfügbar machen.
- Freiheitsgewinn durch KI

Zum KI-Begriff

- Starke KI: wie ein Mensch oder besser
- Schwache KI: Selbst-lernende Systeme zur Lösung spezifischer Probleme, ggf. autonom
- Machine Learning (ML): Algorithmen für (datenbasierte) schwache KI
Der Algorithmus adaptiert sich an Lerndaten (Training, Modellschätzung), ggf. auch im Einsatz
- Rasanter Fortschritt: Umfang Lerndaten, schnelle Computer + Lernalgorithmen
- „Machine learning is statistics. The parameters of statistical models are optimally adapted to data.“ (Bauckhage, Fraunhofer, ‚Trends in Artificial Intelligence‘, 2018).
- ML nutzt oft Big Data (hochdimensional, massiv, schnell)
- Deep Learning (Künstliche Neuronale Netze) orientieren sich am Gehirn

Was ist ein Deep Learner?

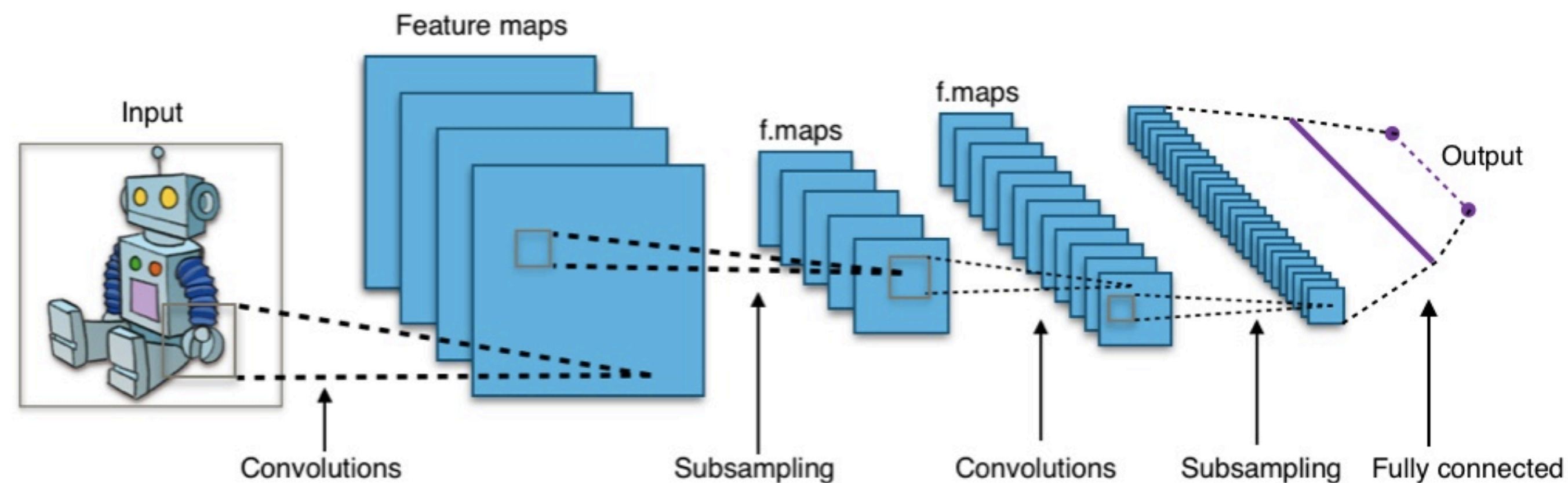
- Wie kann man (mit Schulmathematik) Daten (Punktwolken) klassifizieren?
- Grundstruktur eines künstlichen neuronalen Netzes (KNN)
- Wie kann ein KNN Objekte erkennen, in Beziehung setzen und somit Wissen repräsentieren?
- Wie funktioniert das Erzeugen von KI-Bildern, -Texten usw.?



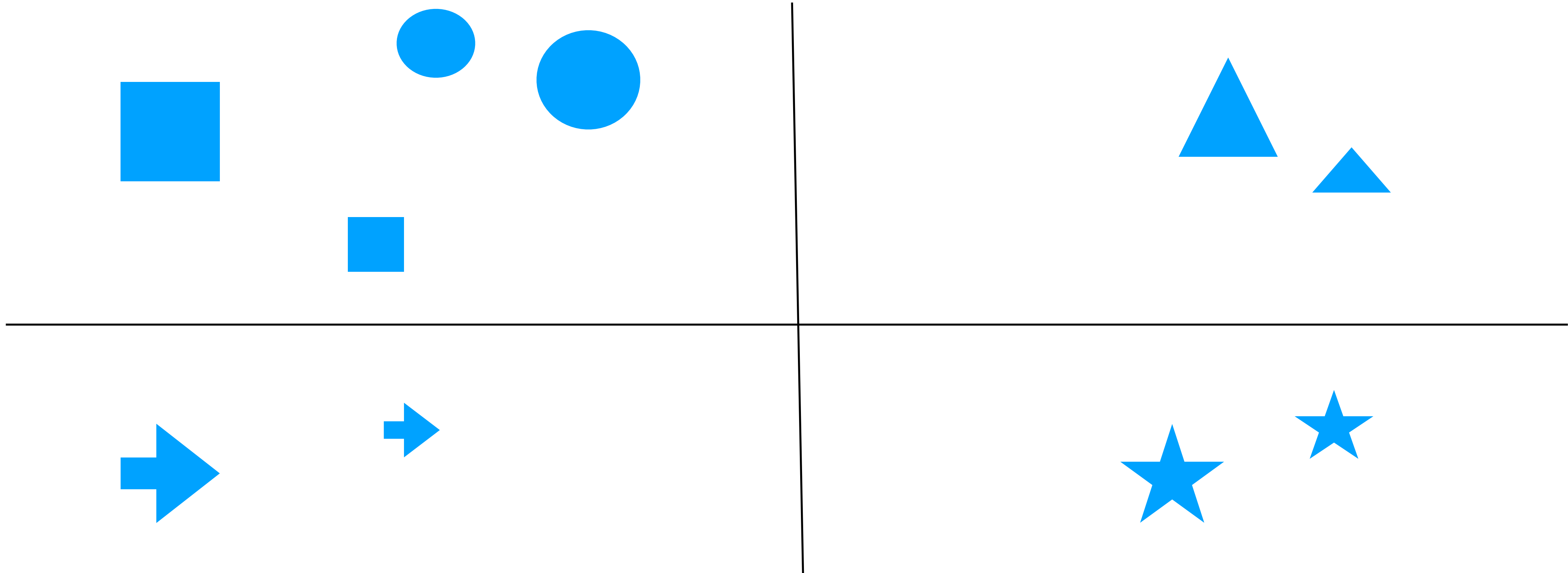
Jedes Neuron berechnet $y = f(w'x)$, x : Eingabe (Inputs oder vorgeschaltete Neuronen), w : Gewichte

$$w'x = \sum_{i=1}^n w_i x_i = w_1 \cdot x_1 + \dots + w_n \cdot x_n \quad (\text{Skalarprodukt der Vektoren } w, x)$$

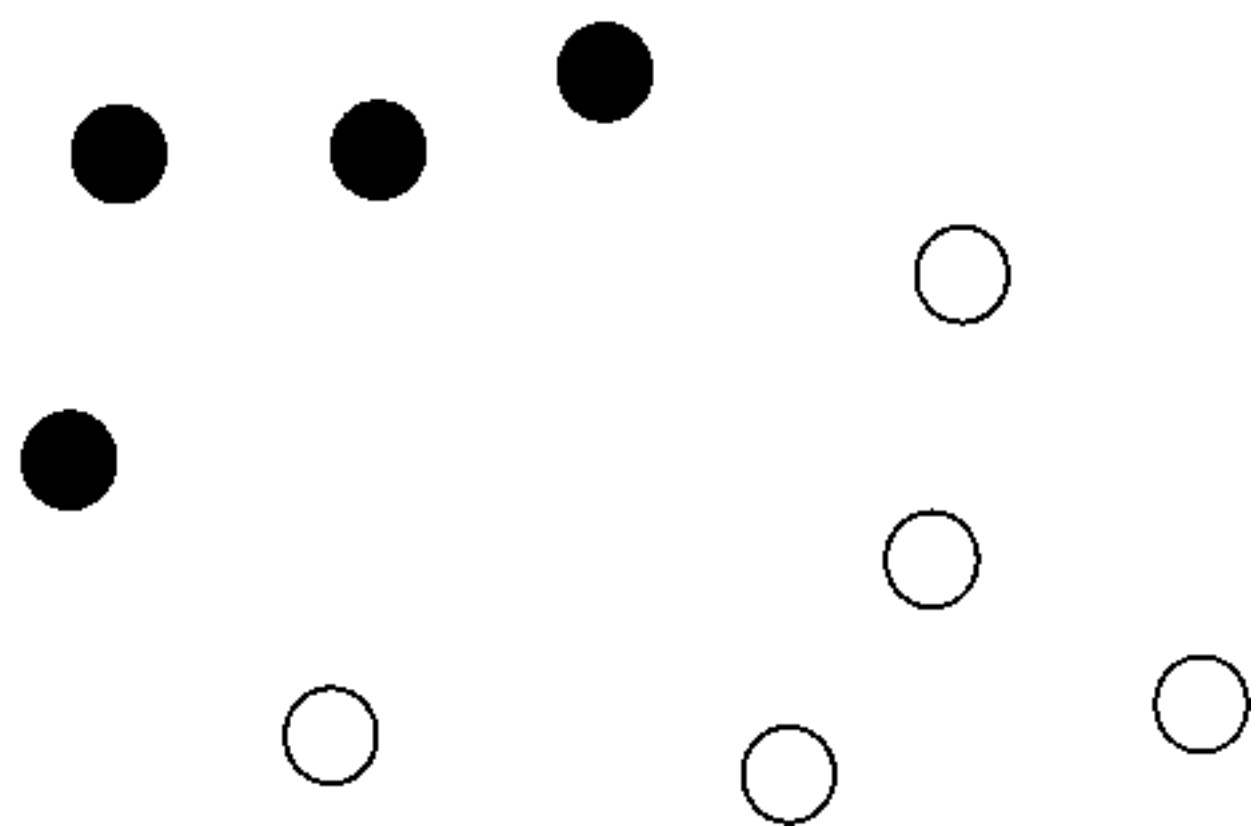
Deep learning: Mehrschichtige neuronale Netze

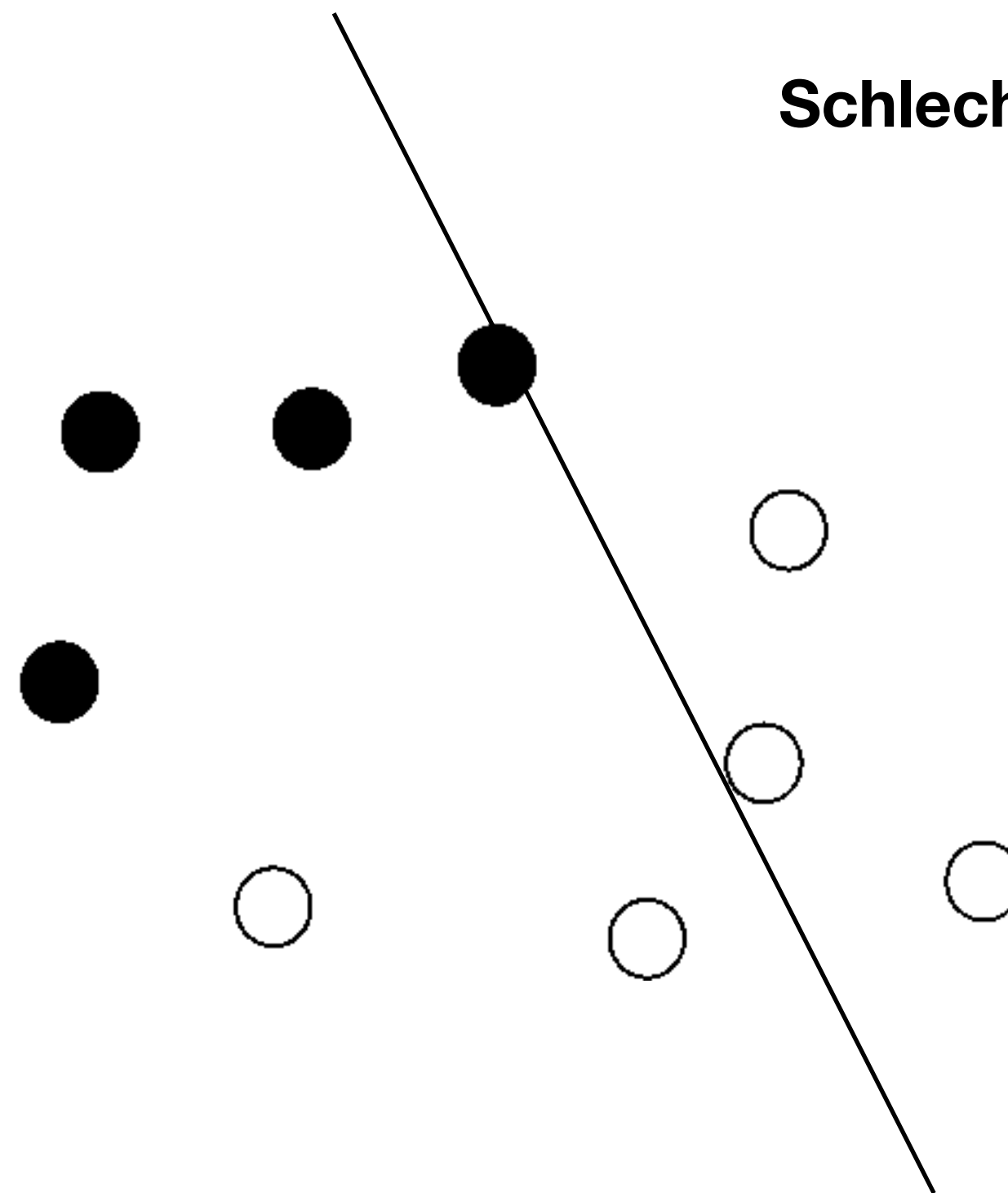


Was ist ein Deep Learner?

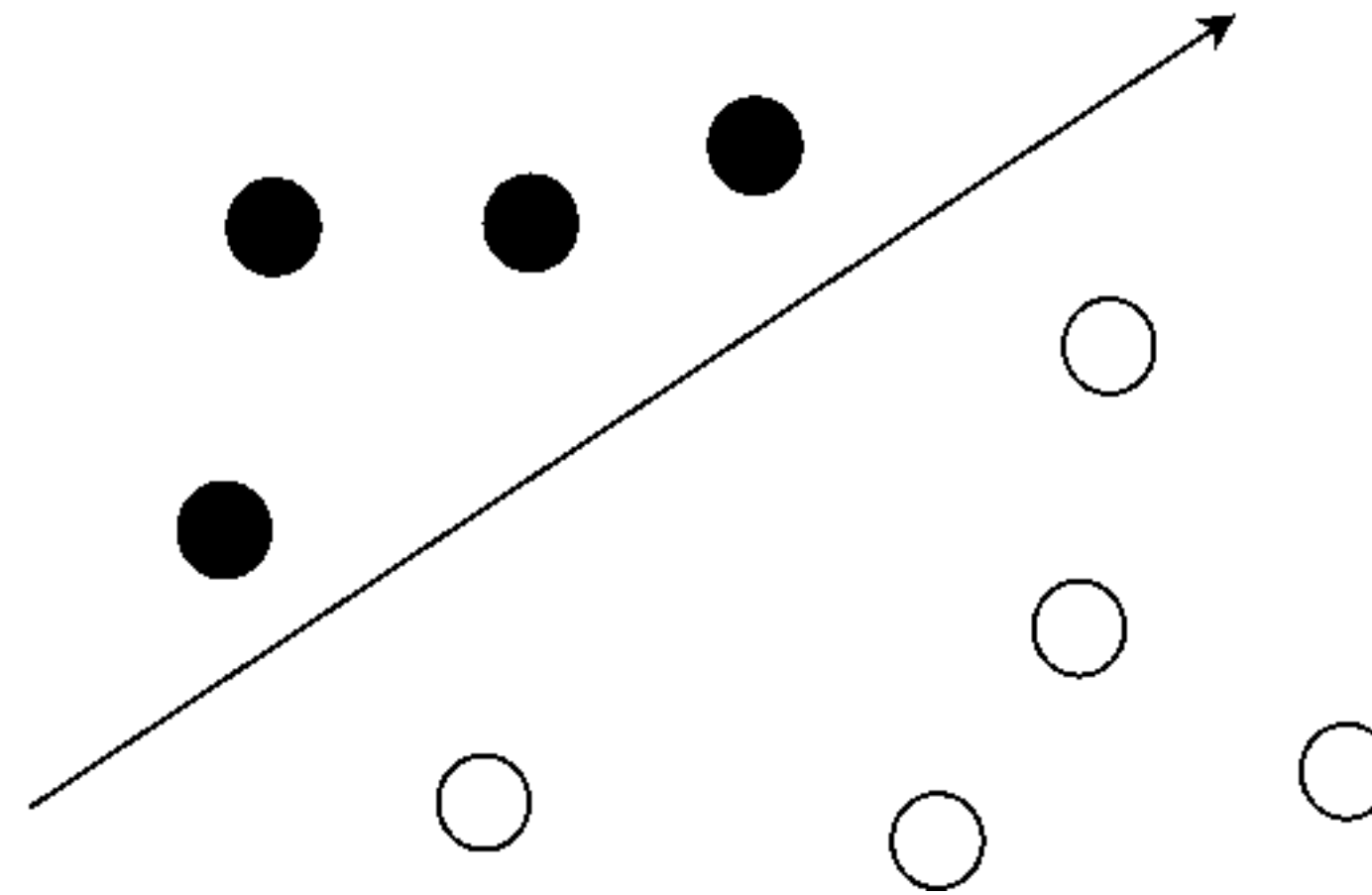


- Wir können diese Objekte recht gut (fast fehlerfrei) durch die zwei Geraden klassifizieren.
- Mit einer dritten Geraden gelingt eine perfekte Klassifikation.



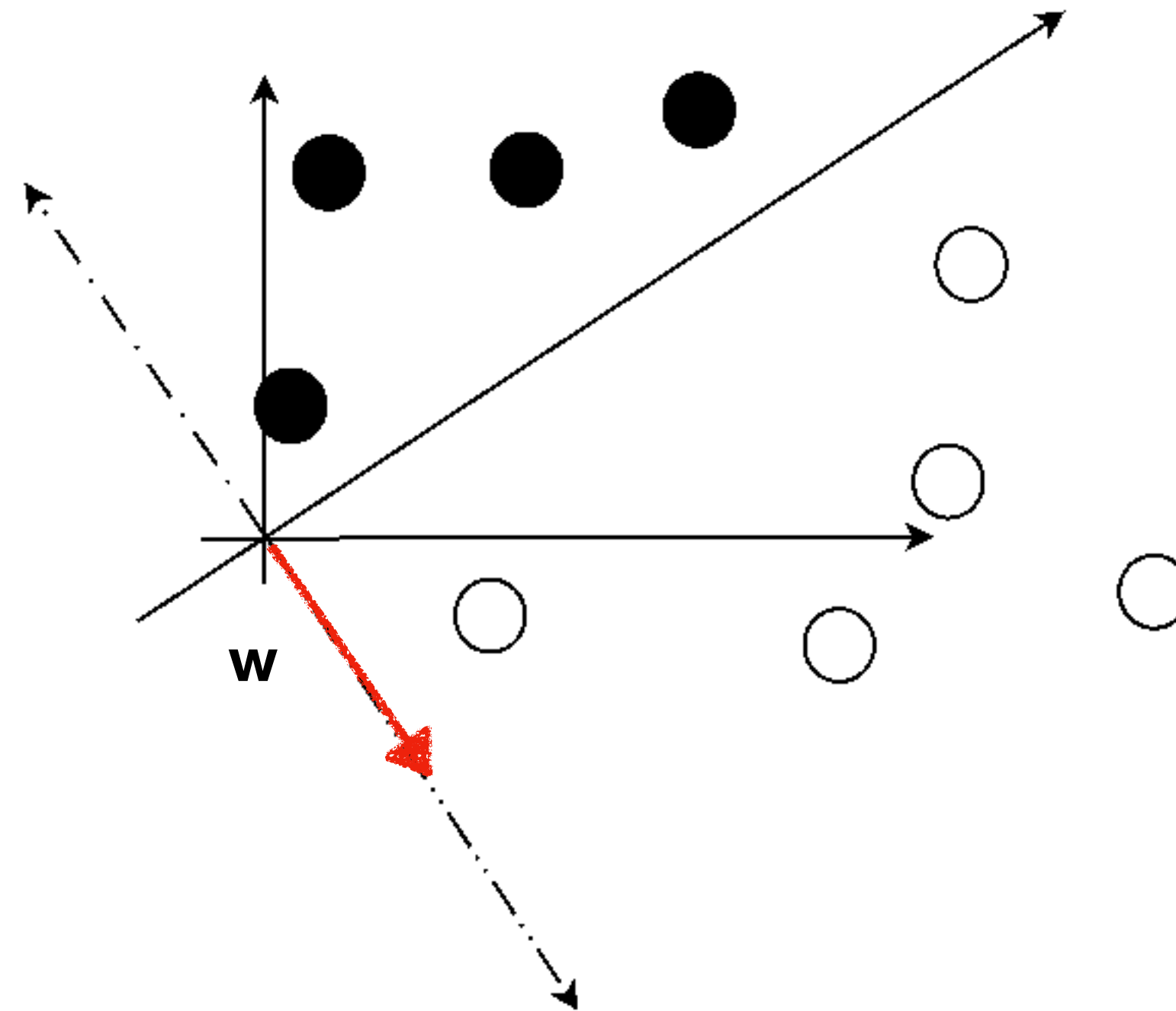


Schlechte Gerade (trennt nicht gut...)

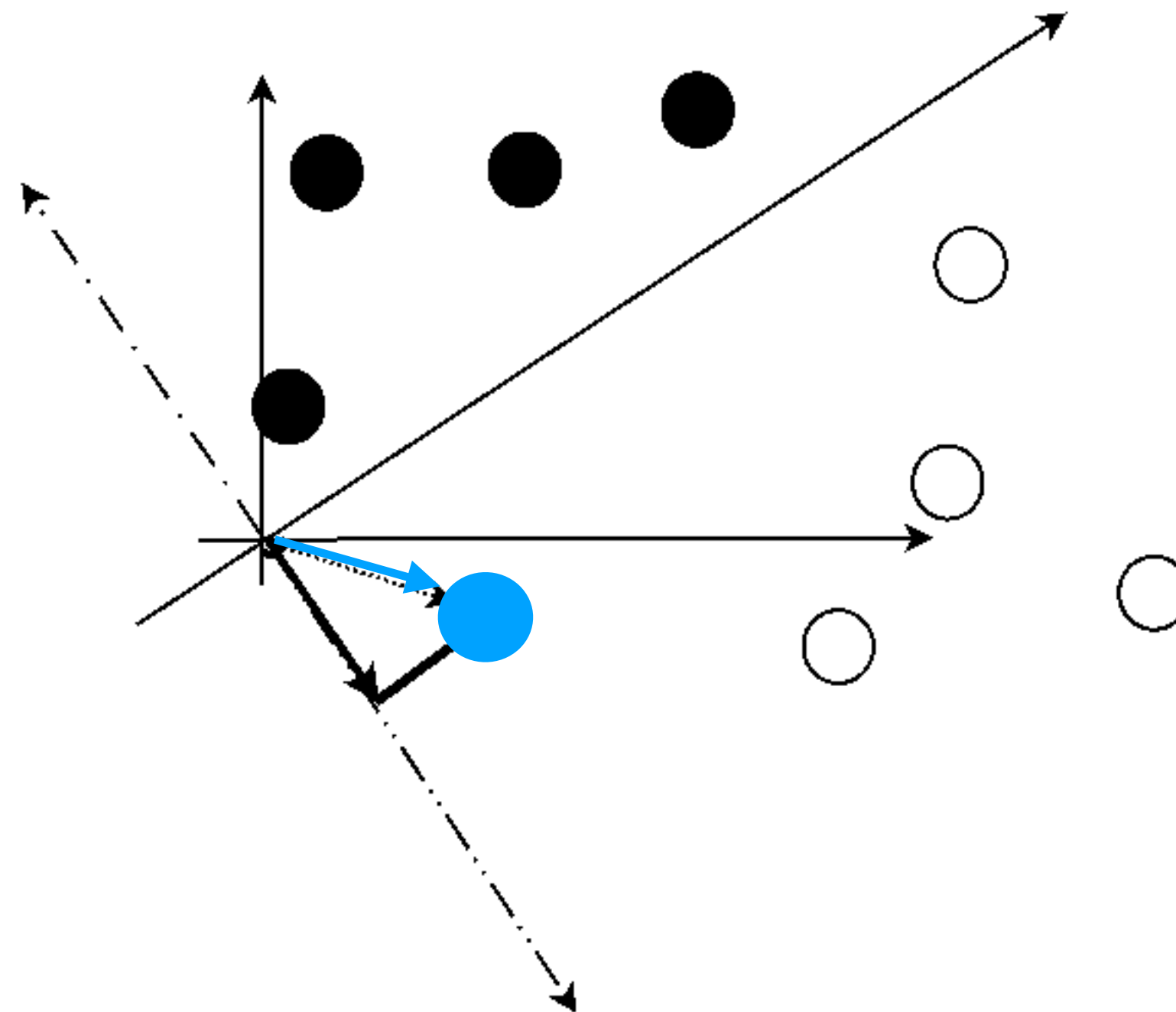


Trennende Gerade G

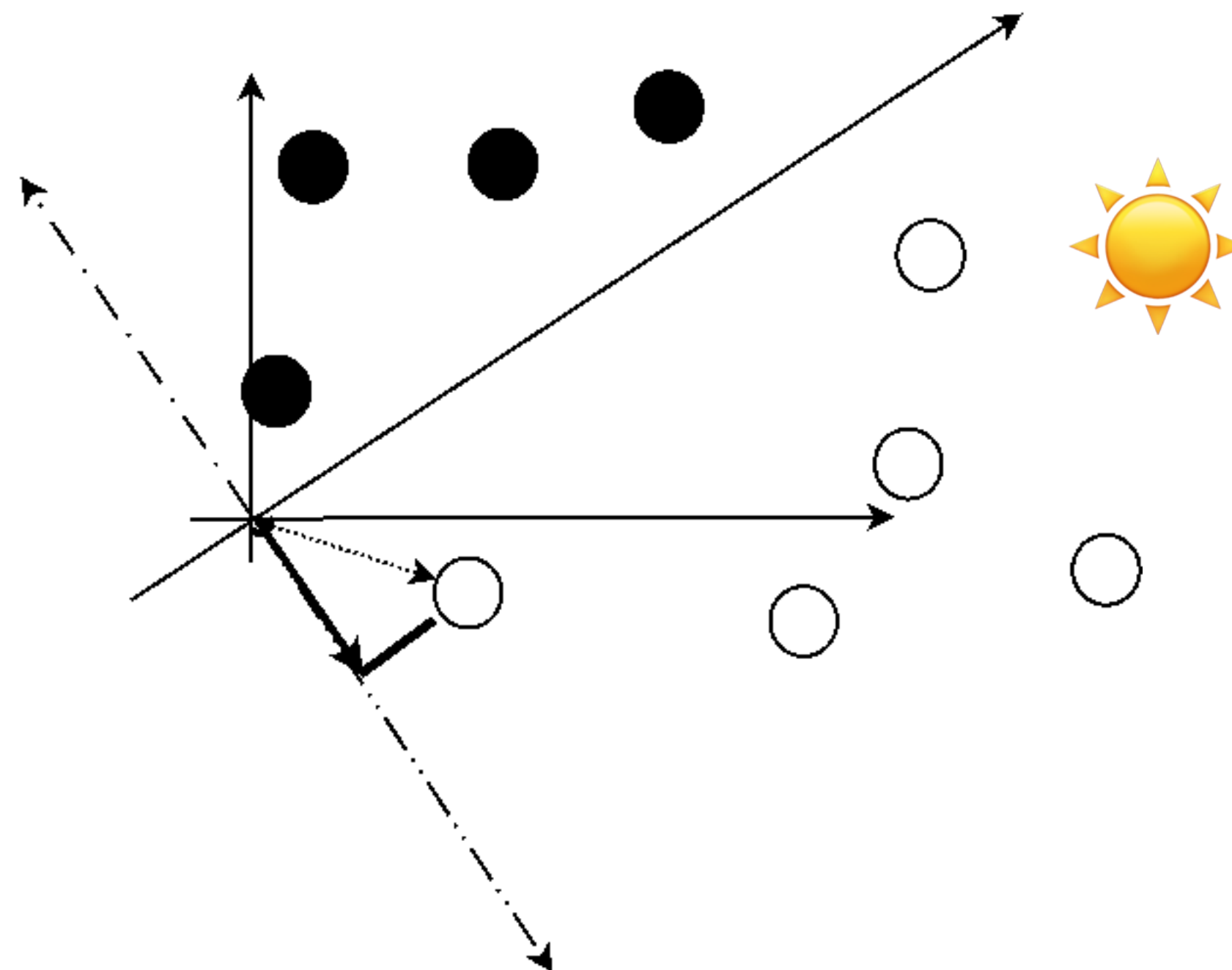
Normale w senkrecht auf G



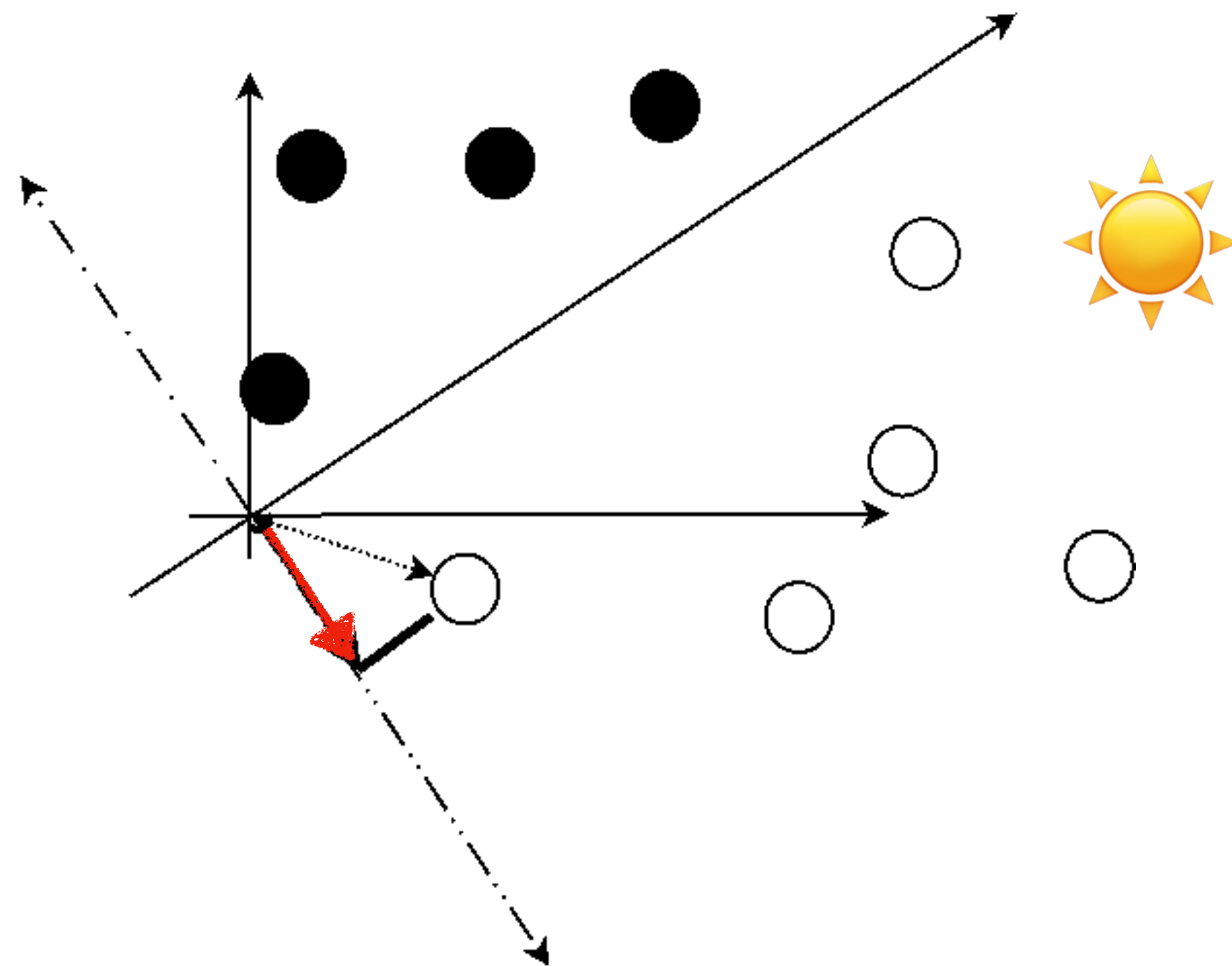
Fälle das Lot



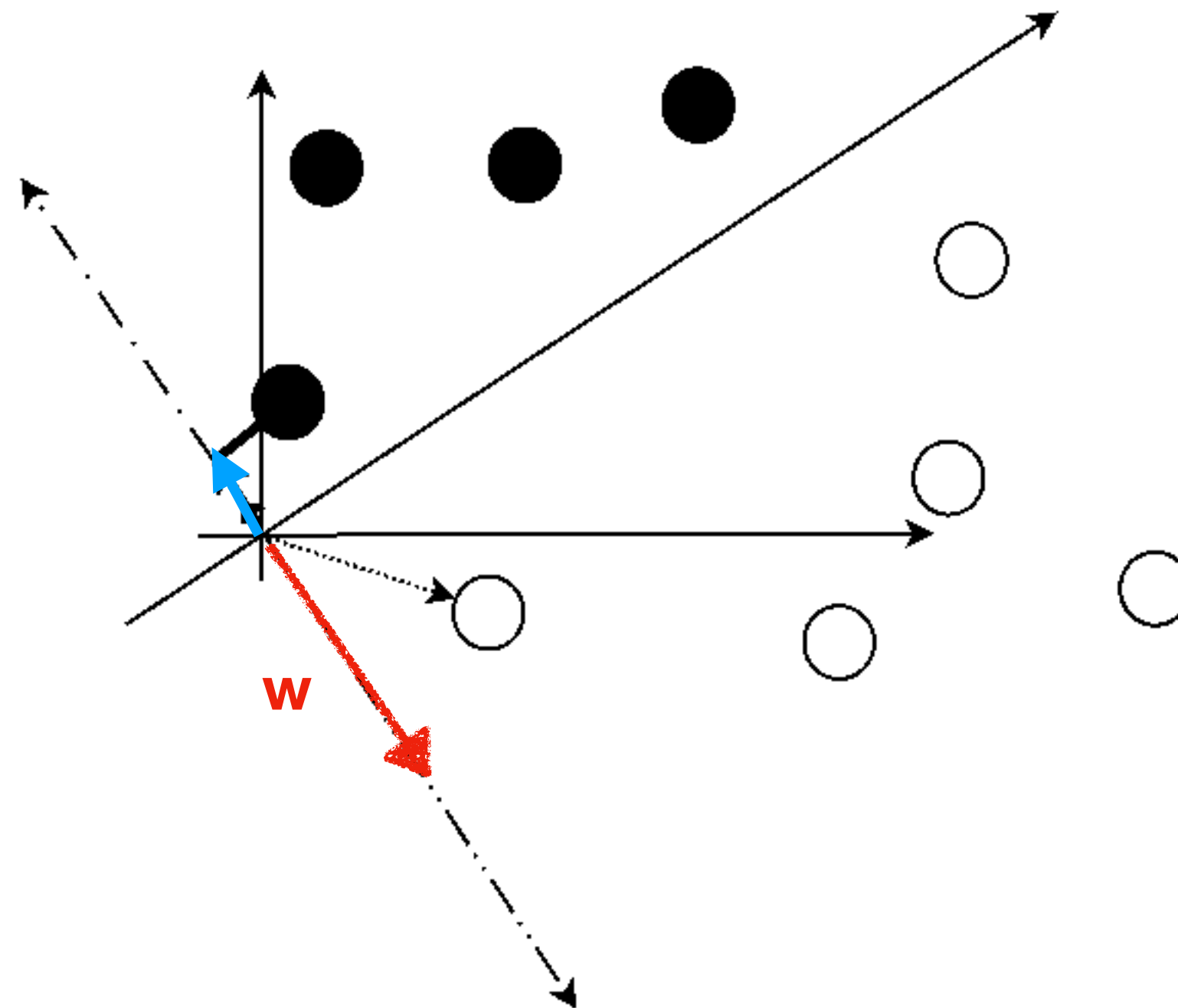
Fälle das Lot



Fälle das Lot



Der rote Vektor zeigt an, auf welcher Seite der Gerade der Punkt liegt!

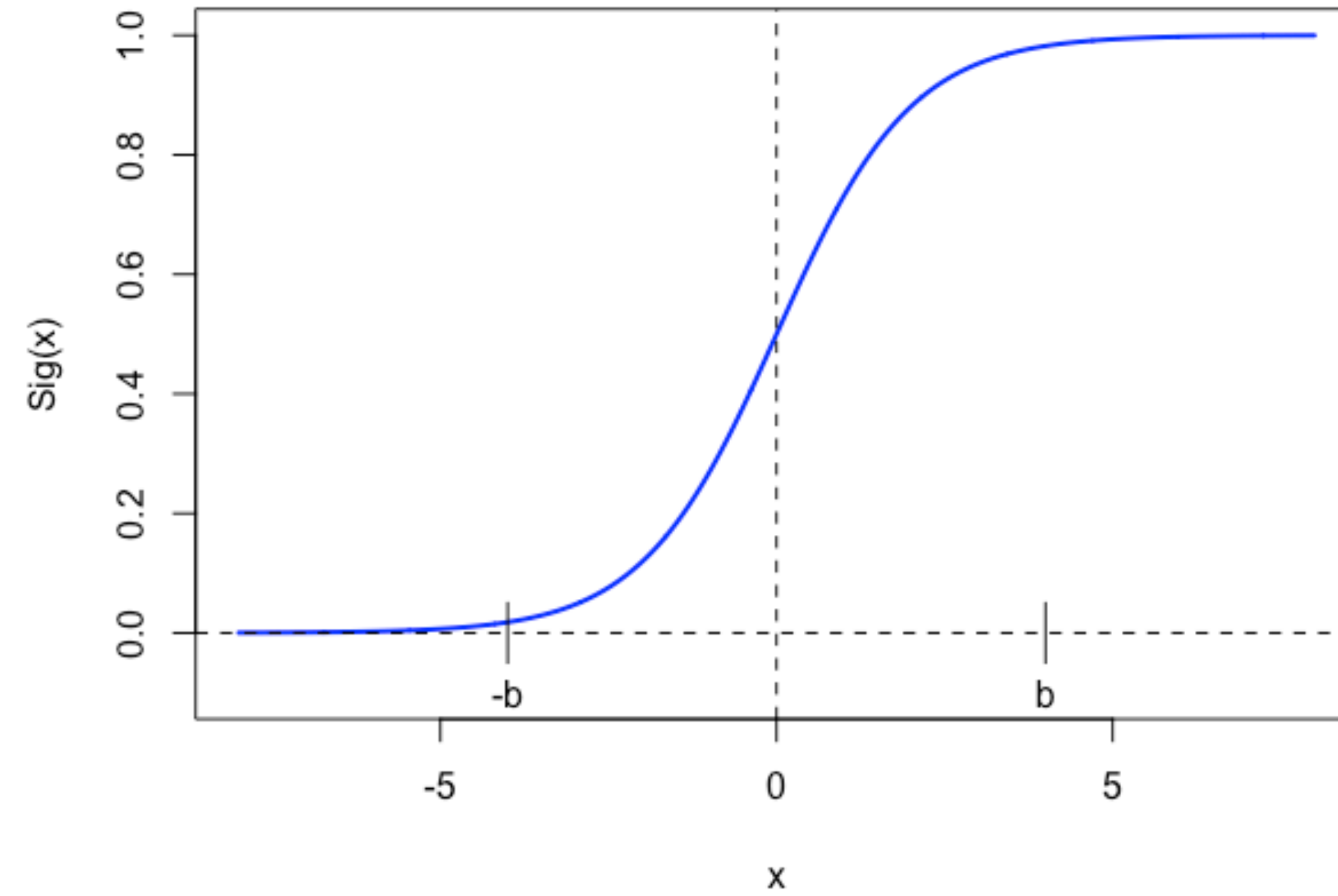


Mathematisch: Skalarprodukt $x'w$ ist die signierte Länge der Projektion auf w , den Normalenvektor der Gerade G

Das Vorzeichen von $x'w$ gibt an, auf welcher Seite der Geraden der Punkt liegt. $|x'w|$ wie weit er weg ist.

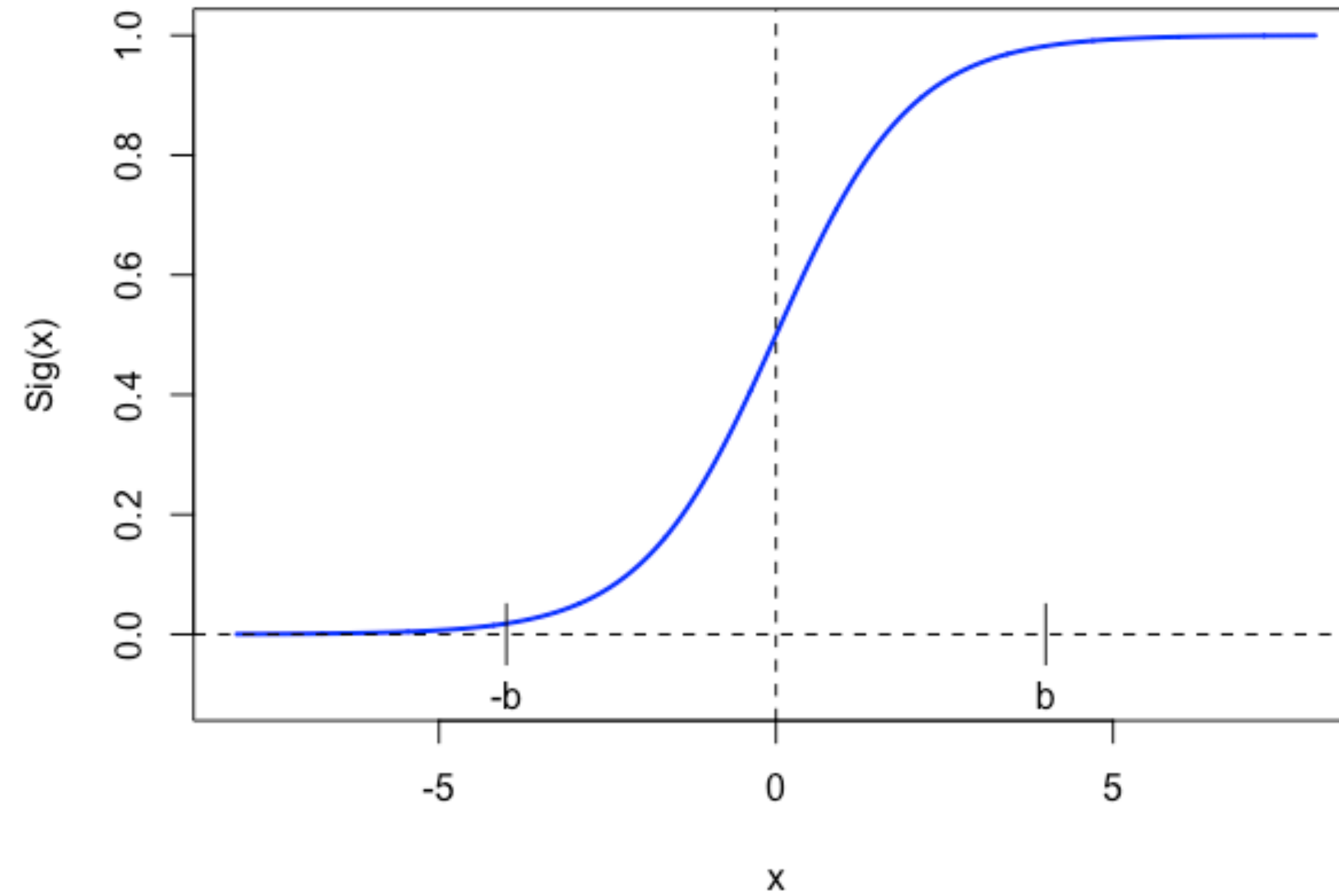
Durch Verändern von w kann man eine gute Gerade suchen, welche die Punktwolke möglichst gut trennt.

Sigmoid-Funktion $S(x)$



Ein künstliches Neuron bewertet nun den Punkt x (Input) durch **$S(b+x'w)$** :

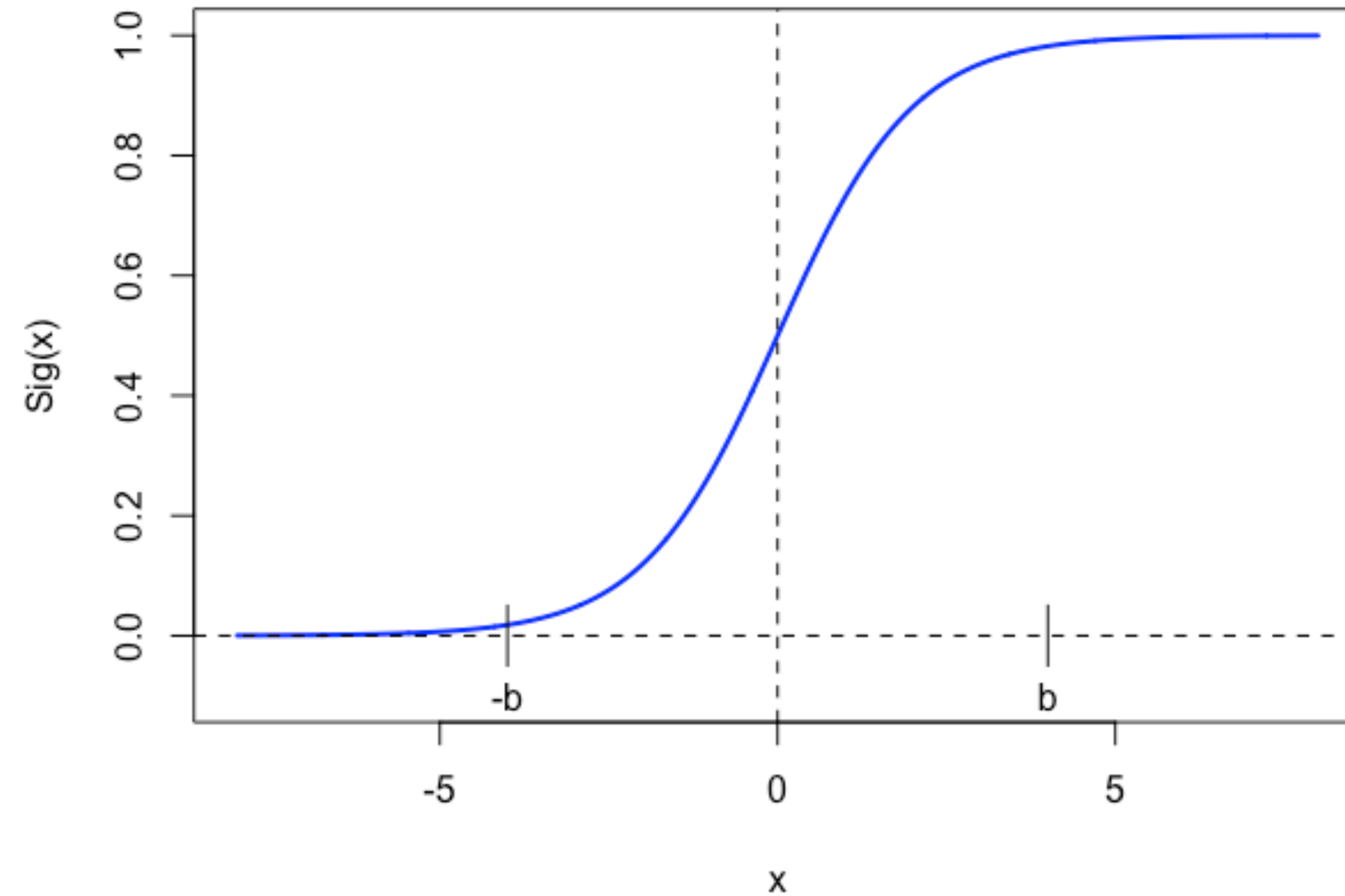
Sigmoid-Funktion $S(x)$



Ein künstliches Neuron bewertet nun den Punkt x (Input) durch $S(b+x'w)$:

- ungefähr 0, wenn der Punkt weit links der Geraden G
- ungefähr 1, wenn der Punkt weit rechts der Geraden G

Sigmoid-Funktion $S(x)$



Ein künstliches Neuron bewertet nun den Punkt x (Input) durch **$S(b+x'w)$** :

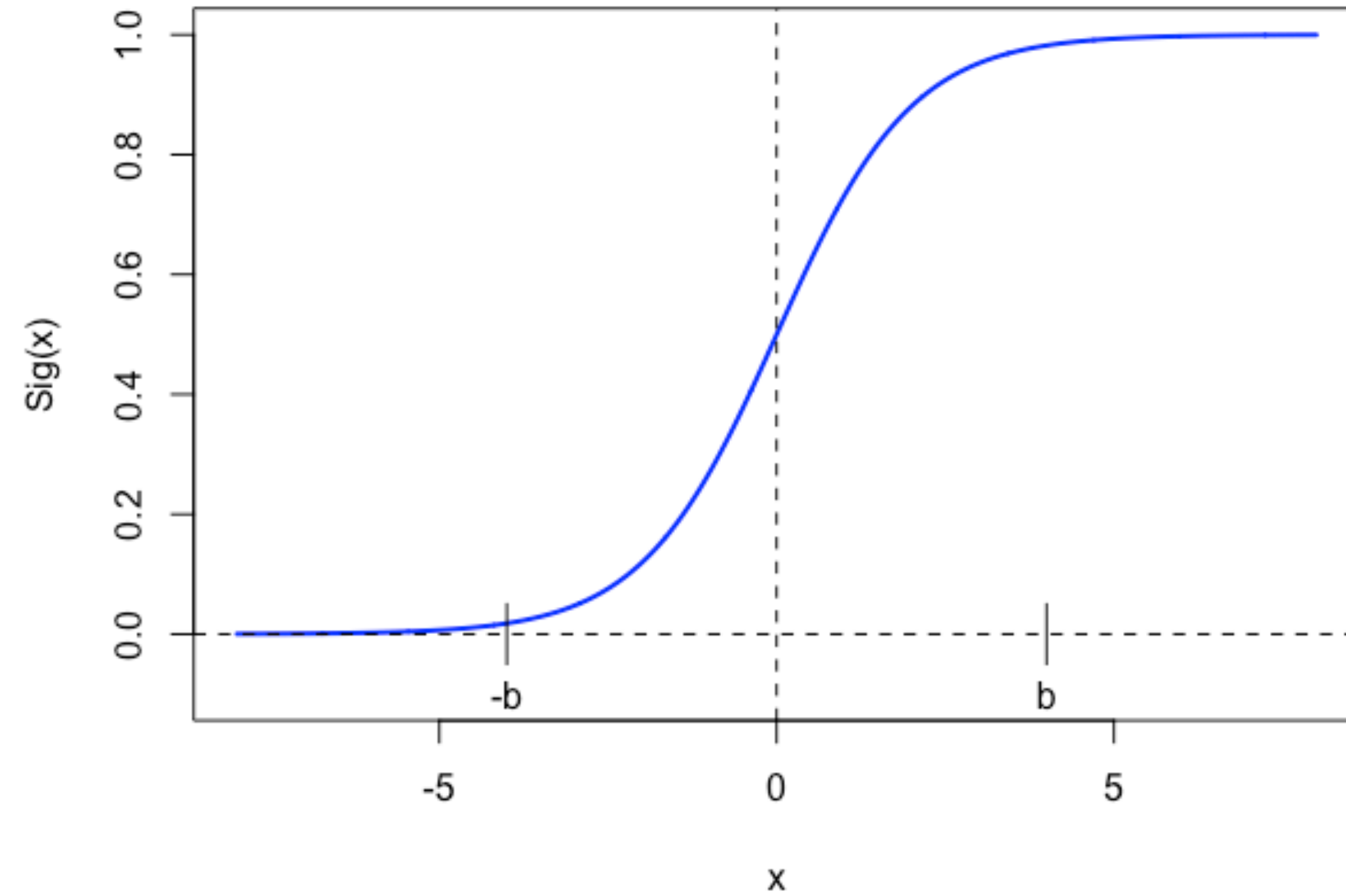
- ungefähr 0, wenn der Punkt weit links der Geraden G
- ungefähr 1, wenn der Punkt weit rechts der Geraden G

Die KI-Algorithmen variieren die Gewichte w, b , um gute Geraden zu finden.

Das Neuron liefert nahezu 0 oder 1, wenn seine Eingabe sehr groß oder sehr klein ist.

Hintereinander geschaltete Neuronen können daher logische Operationen (UND, ODER, ...) implementieren.

Sigmoid-Funktion $S(x)$



$0 * 0 = 0$
 $1 * 0 = 0$
 $0 * 1 = 0$
 $1 * 1 = 1$ (UND)

Ein künstliches Neuron bewertet nun den Punkt x (Input) durch $S(b+x'w)$:

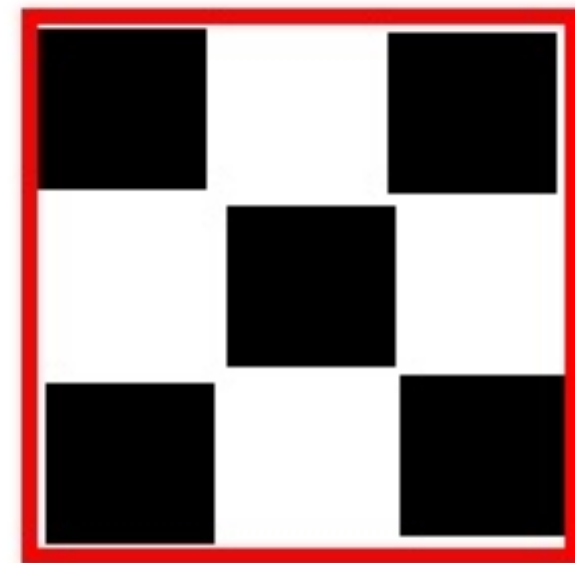
Das Neuron liefert (nahezu) 0 oder 1, wenn seine Eingabe sehr groß oder sehr klein ist.

Hintereinander geschaltete Neuronen können daher logische Operationen (UND, ODER, ...) implementieren.

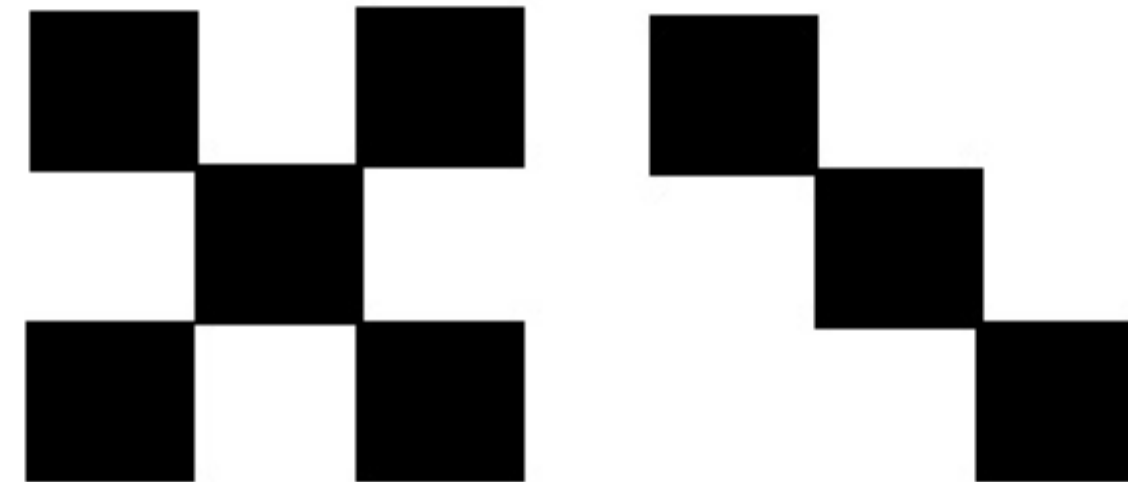
Finden von Mustern (in Fotos)

Schwarz-Weiß-Bild: 1: Punkt gesetzt (schwarz), 0: nicht gesetzt (weiß)

Bildausschnitt:



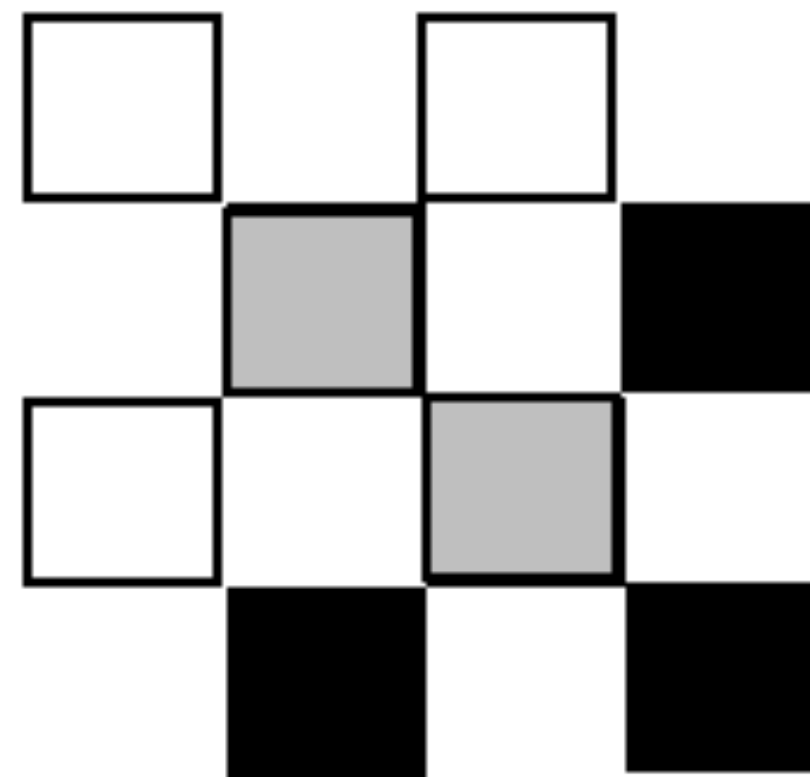
Masken (Zielmuster):



Skalarprodukt: $x \cdot y$ = Anzahl der gesetzten Punkte der Maske,
die einen gesetzten Bildpunkt überdecken.

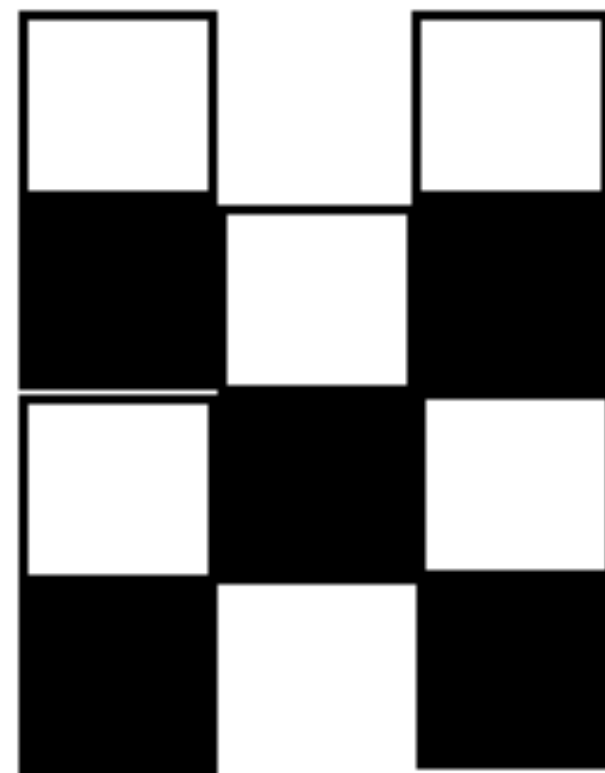
x: Bildausschnitt, y: Maske mit Zielmuster

Finden von Mustern (in Fotos)



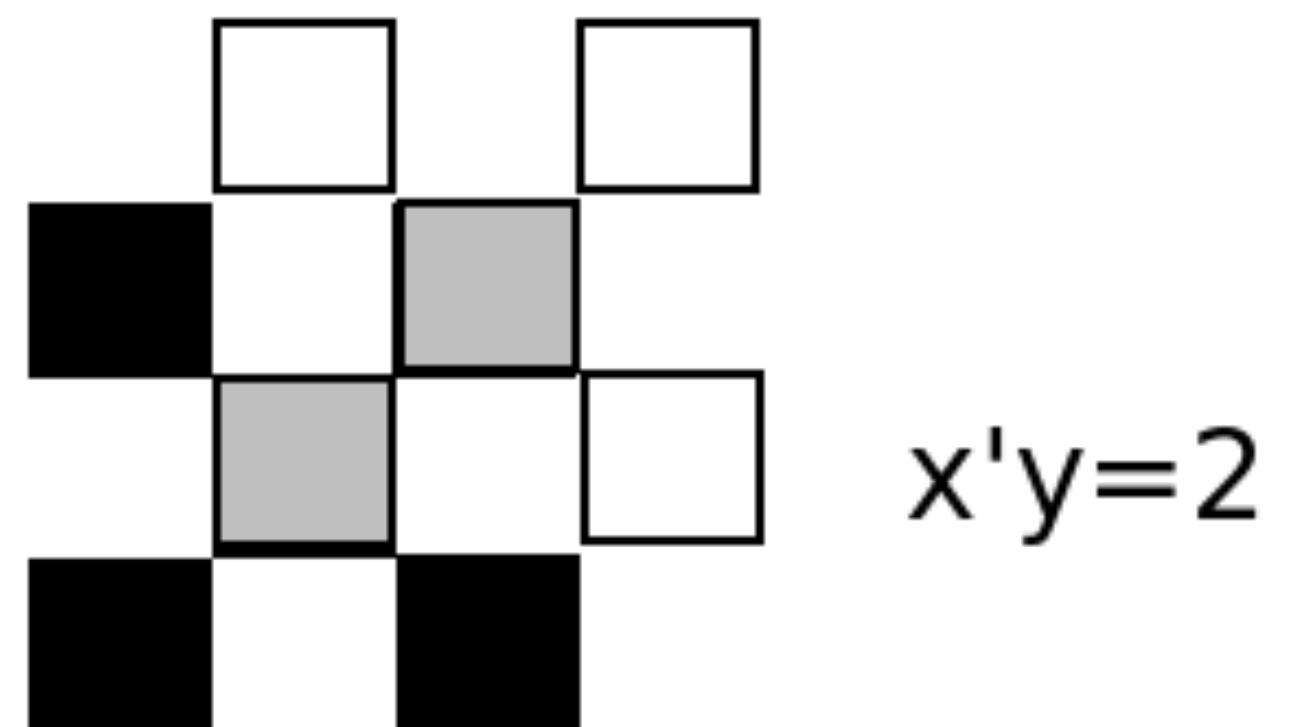
$$x'y=2$$

Finden von Mustern (in Fotos)

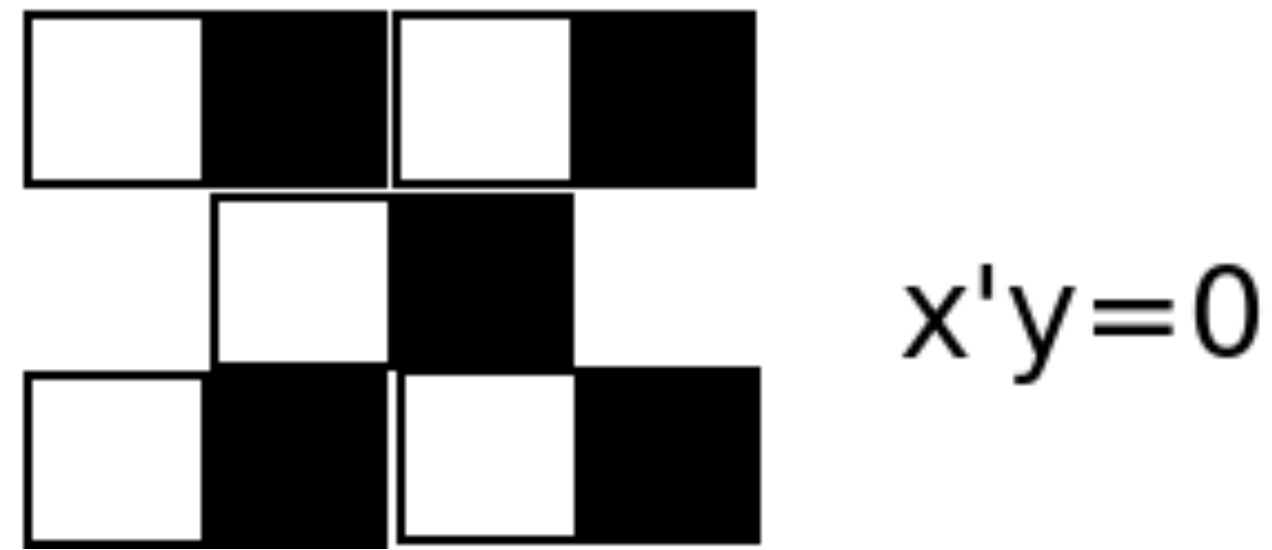


$$x'y=0$$

Finden von Mustern (in Fotos)



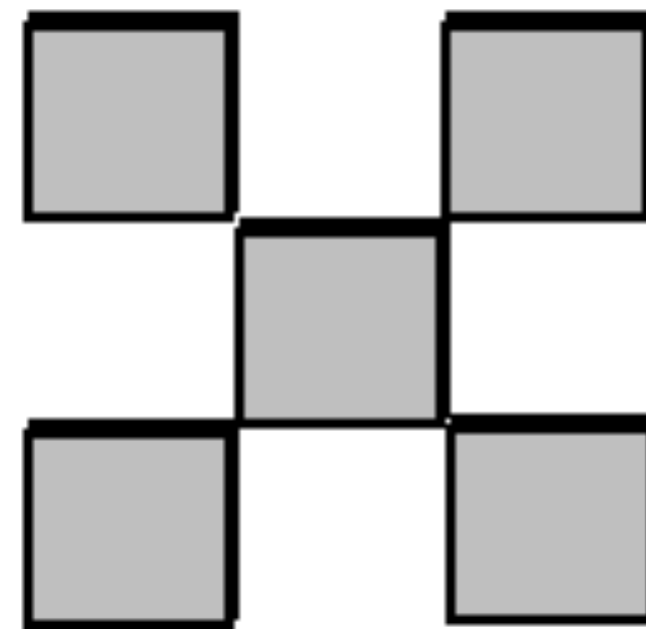
Finden von Mustern (in Fotos)



Nehme für jede mögliche Bildposition des Zielmusters ein Neuron, das durch die $x'y$ -Operation den Match bewertet.

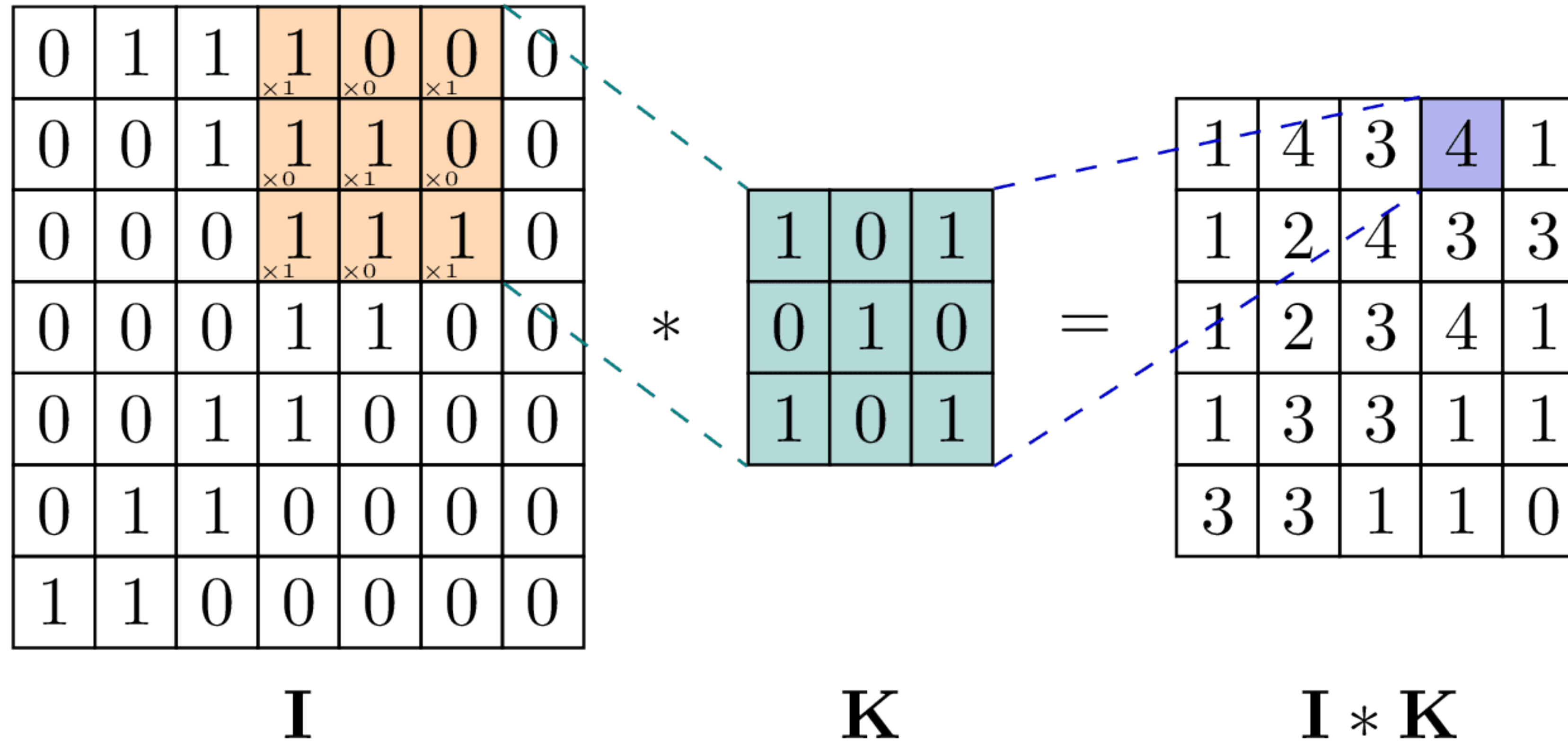
=> Für jedes Zielmuster braucht man

Finden von Mustern (in Fotos)



$$x'y=5$$

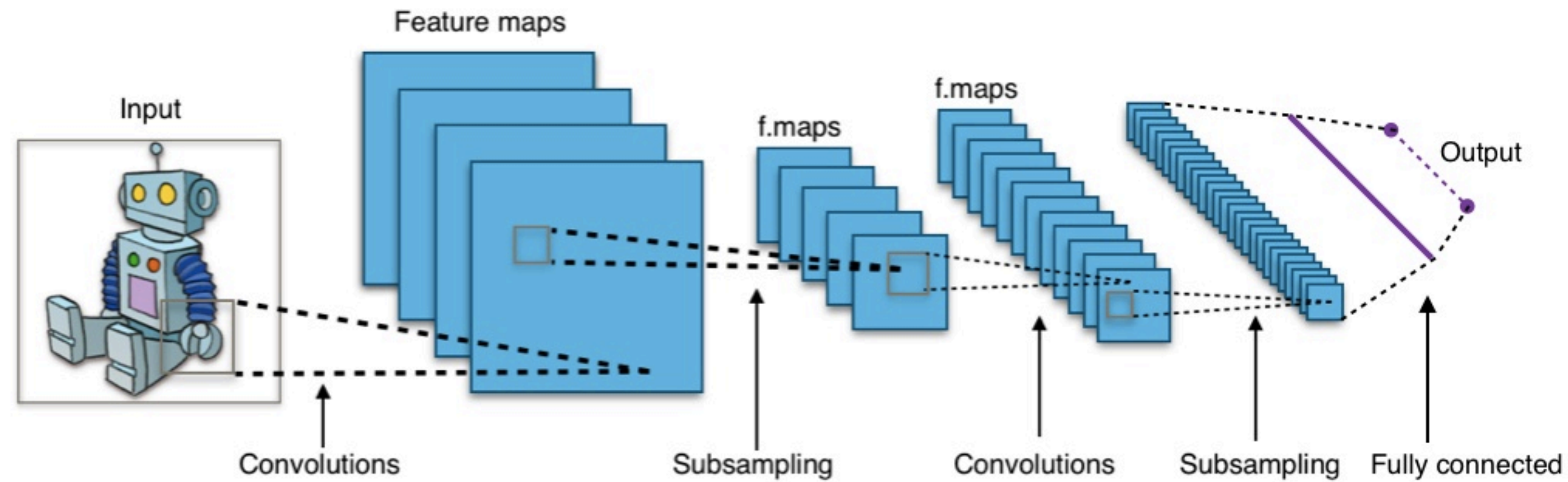
Finden von Mustern...



Nehme für jede mögliche Bildposition des Zielmusters ein Neuron, das durch die x'y-Operation den Match bewertet.

=> Pro Zielmuster für ein NxM Bild braucht man N M Neuronen.

Finden von Mustern



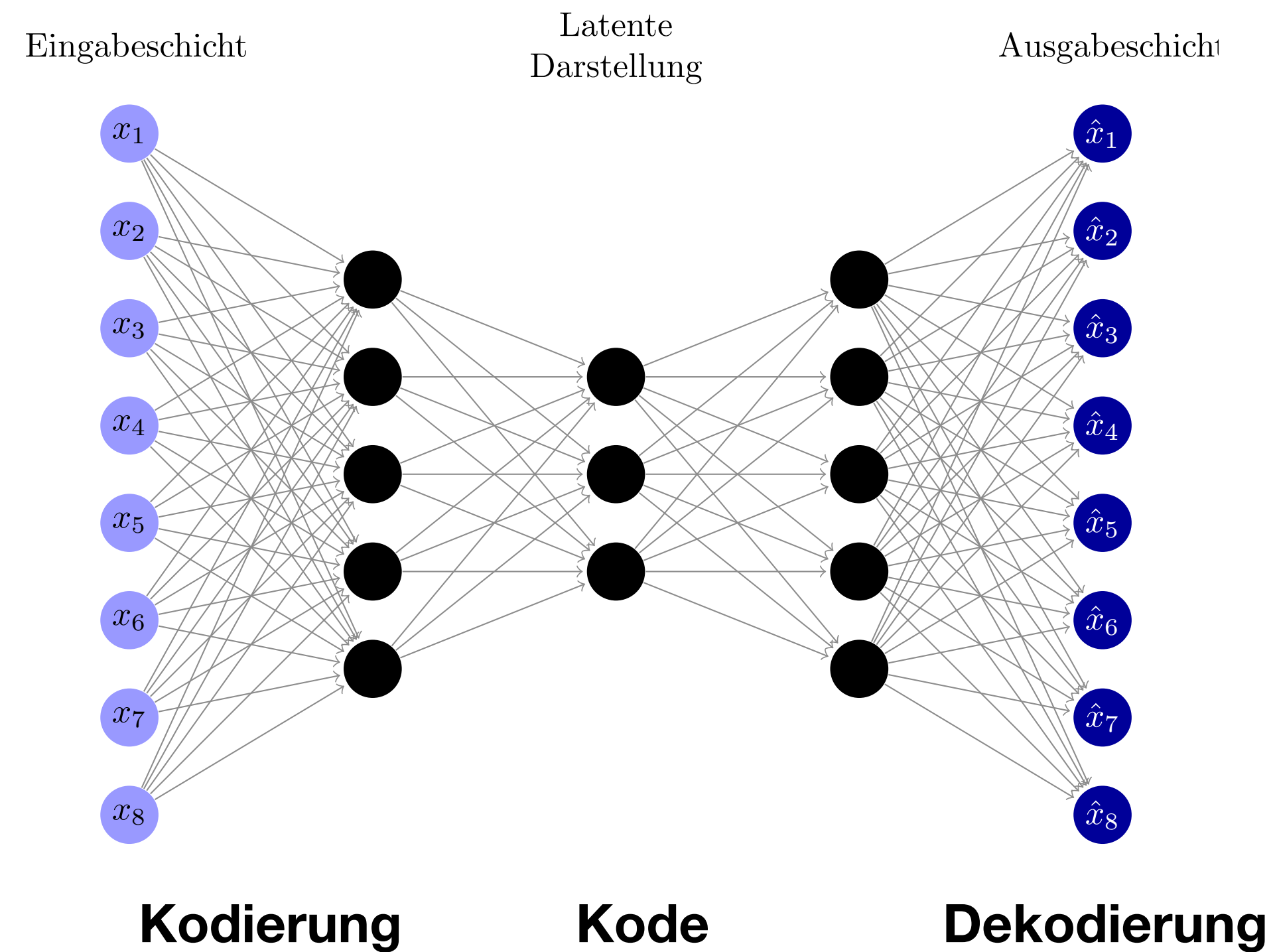
Bei Deep Learning Netzwerke werden die Zielmuster nicht vorgegeben, sondern auch aus den Daten gelernt.

Deep Learning

I _ / | □ △ ○ U ☺ 🕶️ ✈️ 🛟 🧑 -> F -> Freiheit -> Freiheit, Gleichheit, Brüderlichkeit -> „Französische Revolution“

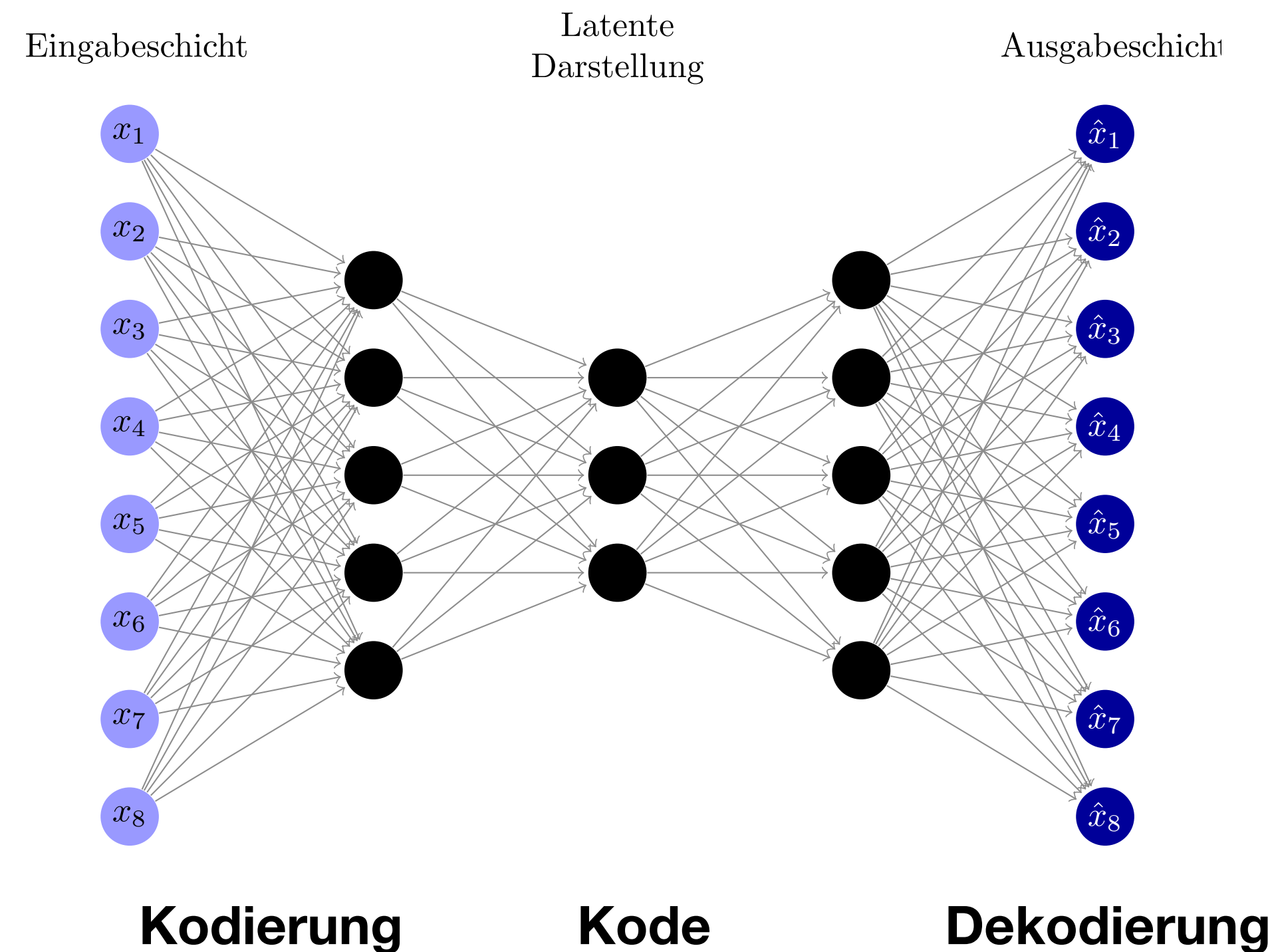
I _ / | □ △ ○ U ☺ 🕶️ ✈️ 🛟 🧑 🧑 -> 🧑 🧑 ✈️ ☀️ 🏖️ 🌃 -> „Paar macht Urlaub unter der Sonne“

Autoencoder-Netzwerke



Engpass (Flaschenhals): Am Engpass muss eine kompakte Darstellung des Gelernten vorliegen.

Autoencoder-Netzwerke



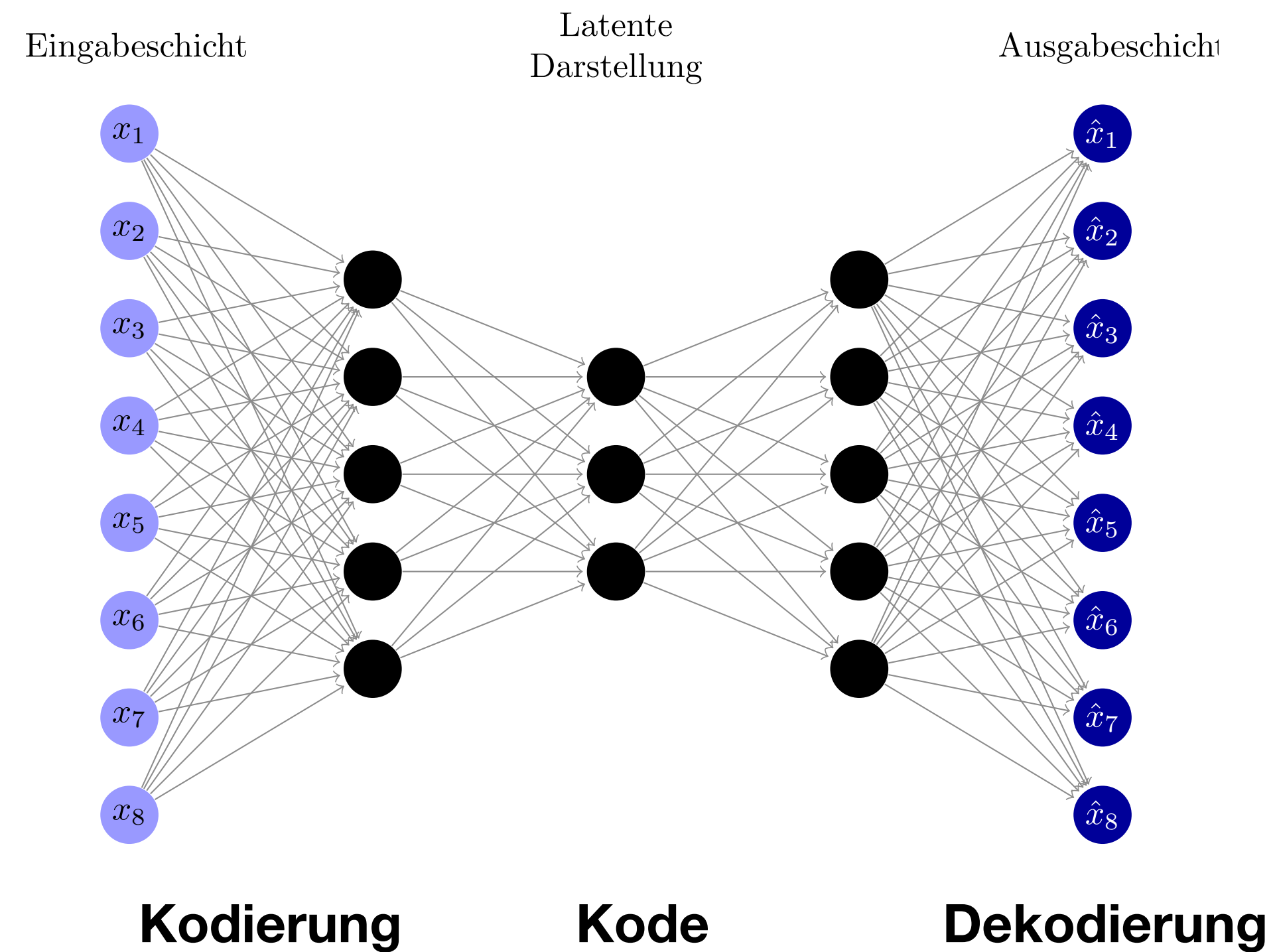
Setze „Input = Output“: Das KNN lernt eine Kode, der bestmöglichst die Daten reproduziert.

Anwendungen: Bilder, Texte,...

Verwendet man nur den zweiten Teil und gibt ein „Kodewort“ vor, so erhält man ein Bild als Ausgabe.

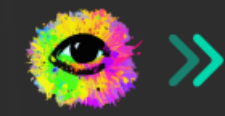
Erweiterung: Gebe Verschlagwortung der Bilder (Kontext) als Input hinzu.

Autoencoder-Netzwerke

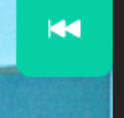
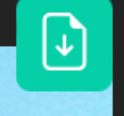


Verwendet man nur den zweiten Teil und gibt ein „Kodewort“ vor, so erhält man ein Bild als Ausgabe.

Erweiterung: Gebe Verschlagwortung der Bilder (Kontext) als Input hinzu.



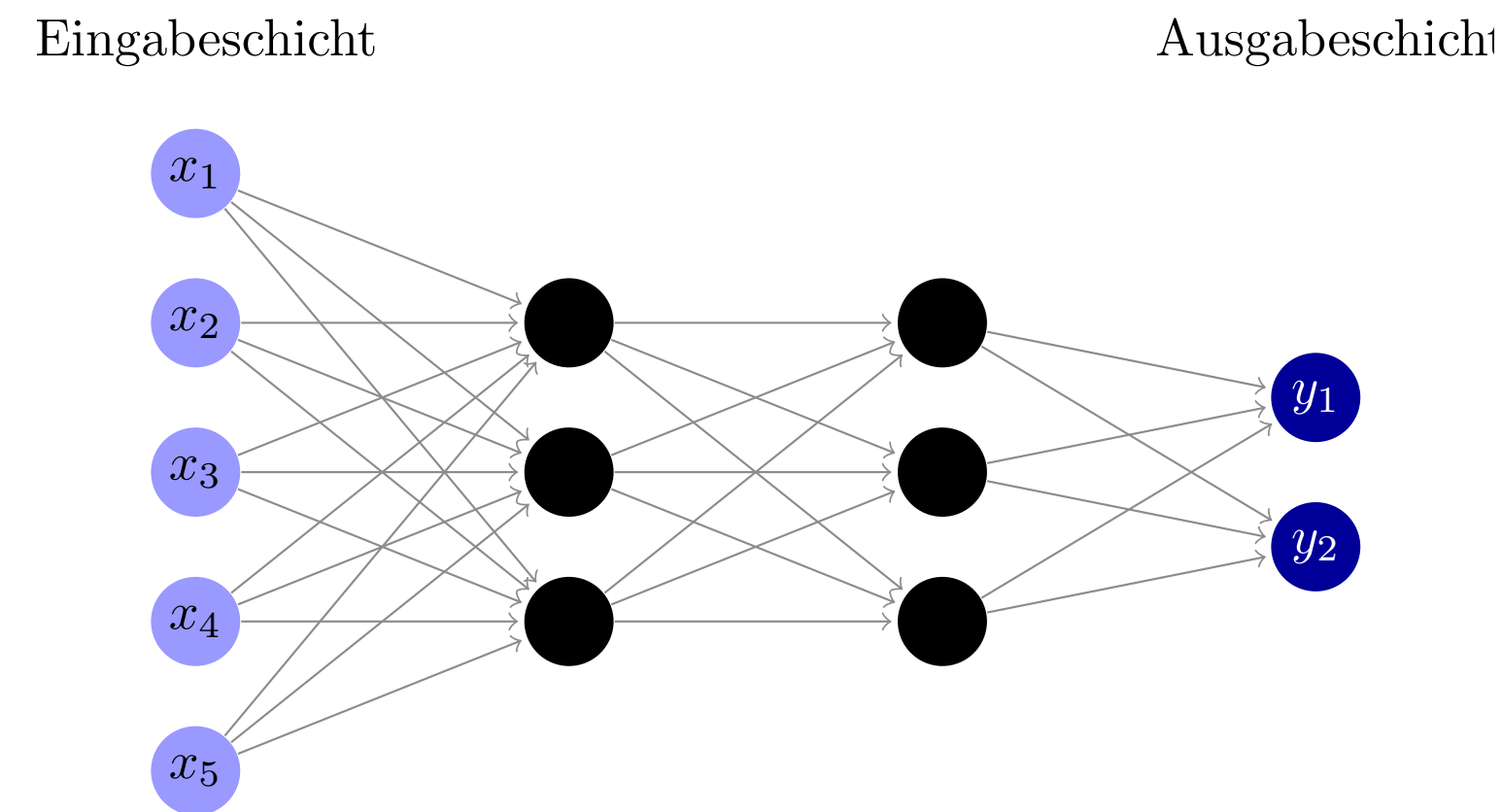
1.0 credits / image



Joe Biden meets Captain Kirk

Dream

Randomisierte Netze: Extreme Learning Machines



**Randomisierte Netze: Wähle die Gewichte der verdeckten Schichten zufällig.
Trainiere das Netz durch Variation der Gewichte der Ausgangsschicht.**

D.h.: Viele zufällige Features werden zur Erklärung der Daten kombiniert.

=> Erhebliche (!) Reduktion des Rechenaufwands

=> Green/Low Cost Machine Learning, ML für wearables, ...

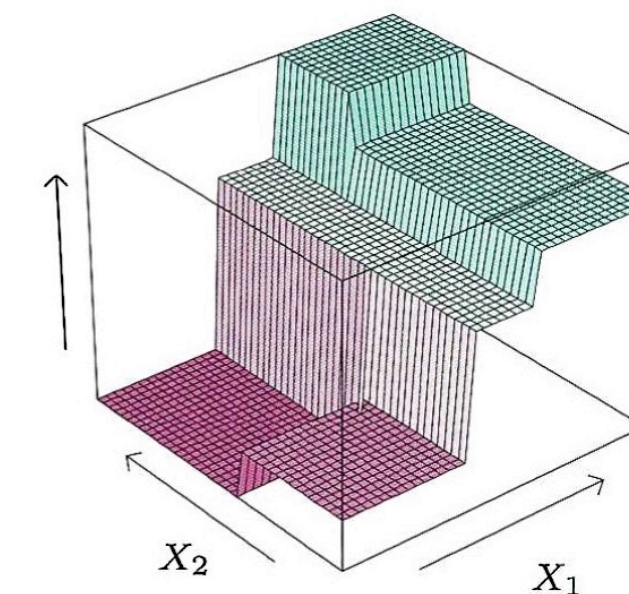
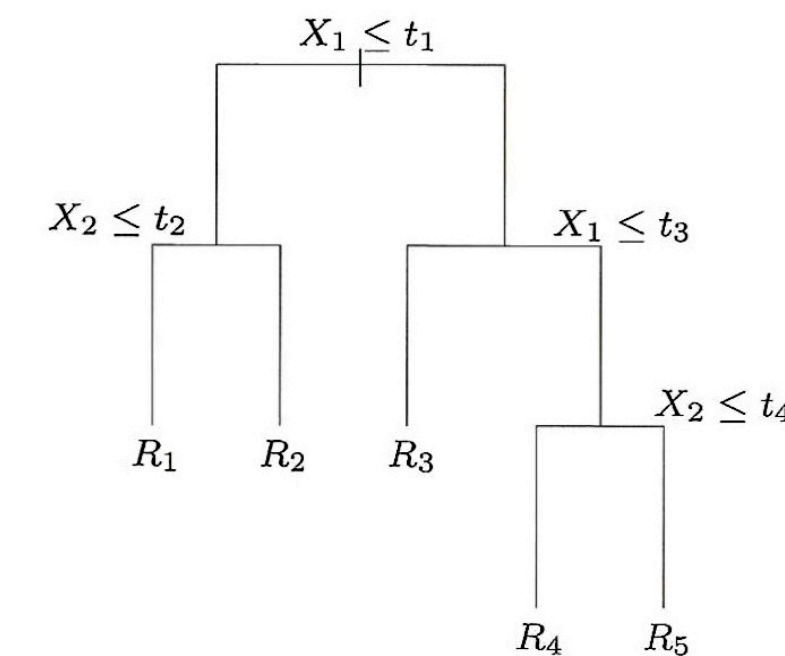
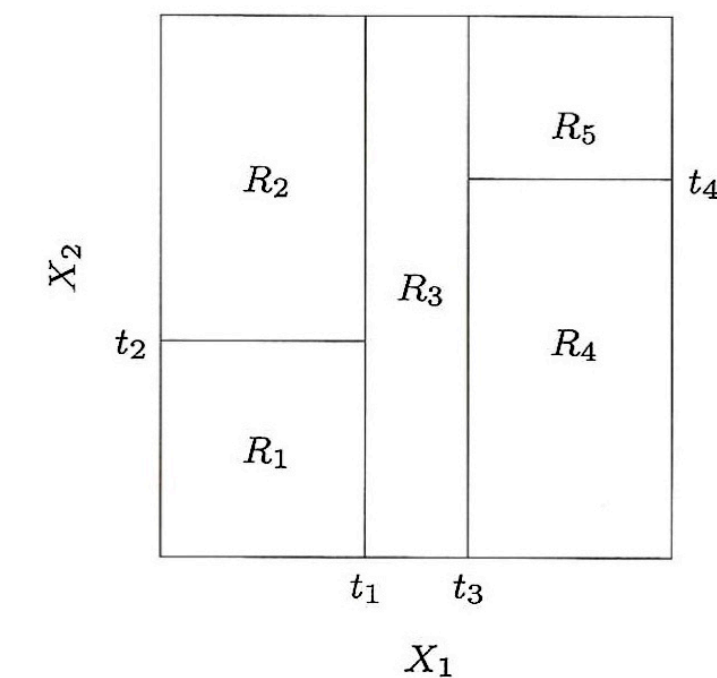
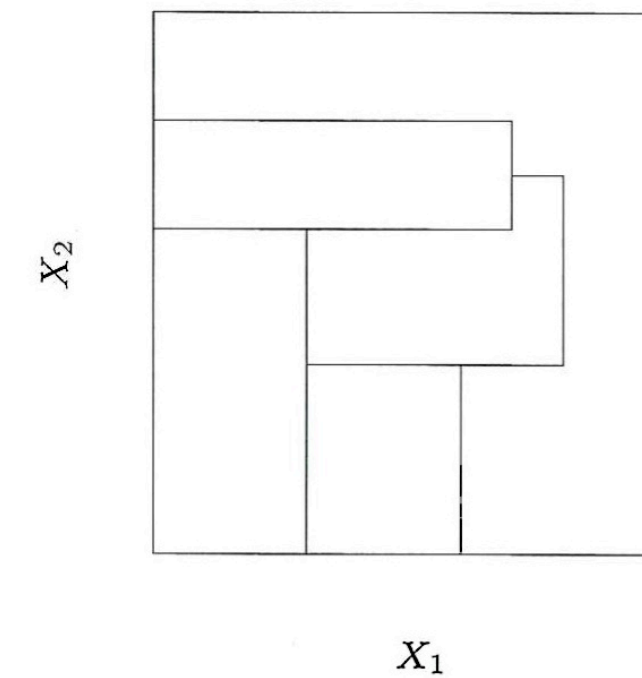
Entscheidungsbäume

X1: BMI, X2: Blutdruck

Entscheidungsbäume schneiden entlang der (oft interpretierbaren) Achsen.

Der Baum wird von oben nach unten konstruiert.

In jedem Schritt wird die Variable gewählt, die den Klassifikationsfehler bestmöglichst verkleinert.



Entscheidungsbäume

X1: Blutdruck, X2: Body-Mass-Index

Entscheidungsbäume schneiden entlang der (oft interpretierbaren) Achsen.

Der Baum wird von oben nach unten konstruiert.

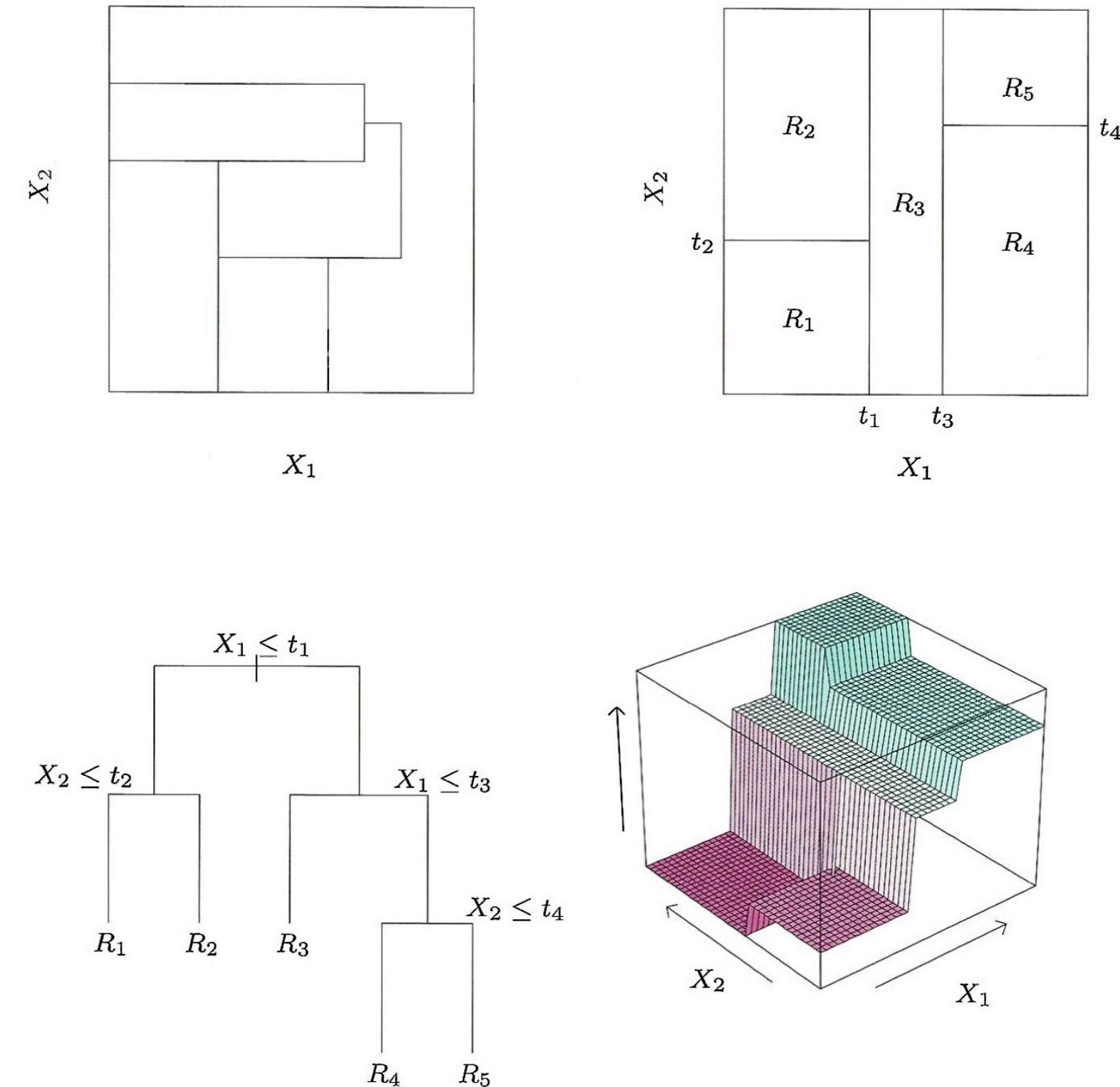
In jedem Schritt wird die Variable gewählt, die den Klassifikationsfehler bestmöglichst verkleinert.

Random Forest:

Zerlege den Datensatz zufällig in M Teile.

**Passe an jeden Teildatensatz einen Entscheidungsbaum an.
(Zufälliger Wald von Entscheidungsbäumen).**

Klassifikation neuer Daten x: Nehme die Mehrheitsentscheidungs der M Bäume.



Vergleichsstudie

Rank	Acc.	κ	Classifier	Rank	Acc.	κ	Classifier
32.9	82.0	63.5	parRF_t (RF)	67.3	77.7	55.6	pda_t (DA)
33.1	82.3	63.6	rf_t (RF)	67.6	78.7	55.2	elm_m (NNET)
36.8	81.8	62.2	svm_C (SVM)	67.6	77.8	54.2	SimpleLogistic_w (LMR)
38.0	81.2	60.1	svmPoly_t (SVM)	69.2	78.3	57.4	MAB_J48_w (BST)
39.4	81.9	62.5	rforest_R (RF)	69.8	78.8	56.7	BG_REPTree_w (BAG)
39.6	82.0	62.0	elm_kernel_m (NNET)	69.8	78.1	55.4	SMO_w (SVM)
40.3	81.4	61.1	svmRadialCost_t (SVM)	70.6	78.3	58.0	MLP_w (NNET)
42.5	81.0	60.0	svmRadial_t (SVM)	71.0	78.8	58.23	BG_RandomTree_w (BAG)
42.9	80.6	61.0	C5.0_t (BST)	71.0	77.1	55.1	mlm_R (GLM)
44.1	79.4	60.5	avNNet_t (NNET)	71.0	77.8	56.2	BG_J48_w (BAG)
45.5	79.5	61.0	nnet_t (NNET)	72.0	75.7	52.6	rbf_t (NNET)

Studie: Vergleich von 197 Klassifizierern über 121 Datensätze (UCI Repository)

- 1. Random Forest (parRF_t)**
- 2. Support Vector Machines (SVM)**
- 3. Extreme Learning Machines (elm_kernel_m)**

Aktuelle relevante KI am Bsp. GPT-3

- Relevante KI-Infrastruktur und KI-Modelle wie GPT-3 werden von Privatunternehmen entwickelt (OpenAI, Amazon, Google, Tesla, ...)
- GPT-3: 96 Schichten, 175 Mrd. Parameter, Daten: ca. 500 Mrd. Token Text
MoE`21 (Google): 36 Schichten, 600 Mrd. Parameter
GLaM (Google): 36 Schichten mit jew. 64 Experten, 1,2 Billionen Par., Daten: 1,6 Billionen Token Text

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

- Energiebedarf für 1xTraining:
 - Tesla: ca. 1 Supercomputer für 1 Tag
 - GPT-3: 1000 Mwh (Energiebedarf einer Kleinstadt pro Tag), Gehirn: ca. 20W
- Konsequenzen:
 - Wissenschaftler können nicht auf dieser Skala Modelle entwickeln und (empirisch) untersuchen.

Einige Trends

- **Imitation Learning:** KI soll durch Abgucken vom Menschen lernen.
OpenAI: KI trainiert mit 70000h Minecraft Videos.
- **Multimodale KI:** Kombination von Audio, Text, Videos.
- **NeuroAI:** Orientierung an Fortschritten der Neurowissenschaften,
Einbezug von Sensorik (Zador et al.)
- Brauchen wir ein „AI-CERN“ und wird Open-Source (GP)AI überleben?
Große KI wird von Privatunternehmen entwickelt.
Aufwand großer KI-Modelle sprengt akademischen Rahmen.

II. Fallstricke

Fallstricke: Datenbasis

- Datenqualität, Datenherkunft, Datenkuration.
- Erhebungsdesign: Zufallsstichprobe versus gegebene Datenbasis
- Zusammenfassung von Datensätzen zulässig?
- (Selection) Bias
„Lernen“ aus Internet-Daten (z.B. soziale Netzwerke) - Aber was lernt man da und sollte man das alles ‚lernen‘?
- Analysen und Schlüsse aus Daten erfordern valide Datenbasis
- Statistik liefert Grundvokabular/Grundwissen (Grundgesamtheit, Stichprobe,...)

Fallstricke: Simpson's Paradoxon

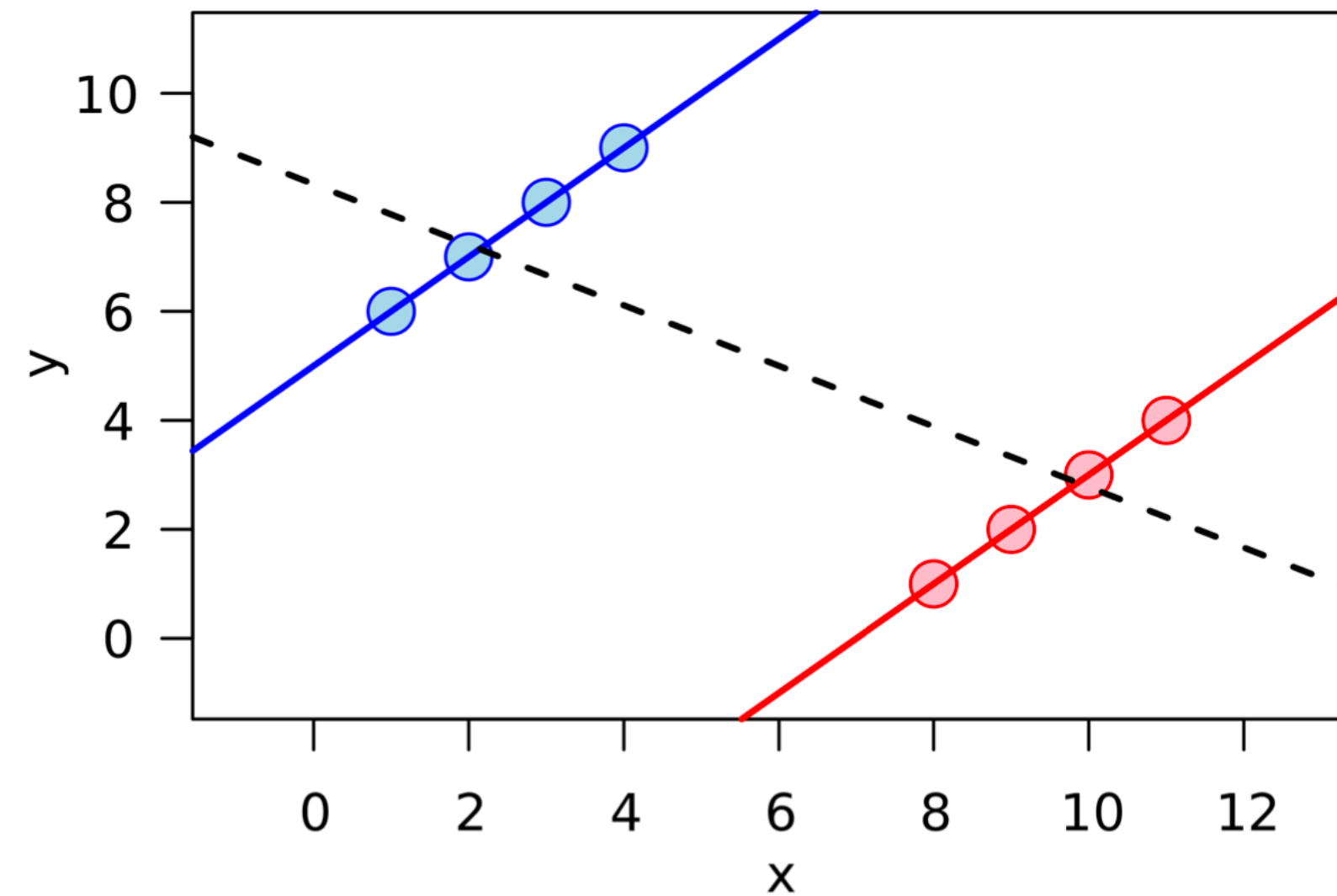


Fig. 2 Simpson's paradox for continuous data: a positive trend is visible for both groups individually (red and blue), but a negative trend (dashed line) appears when the data are pooled across groups (Wikipedia, 2020)

Fallstricke: Simpson's Paradoxon

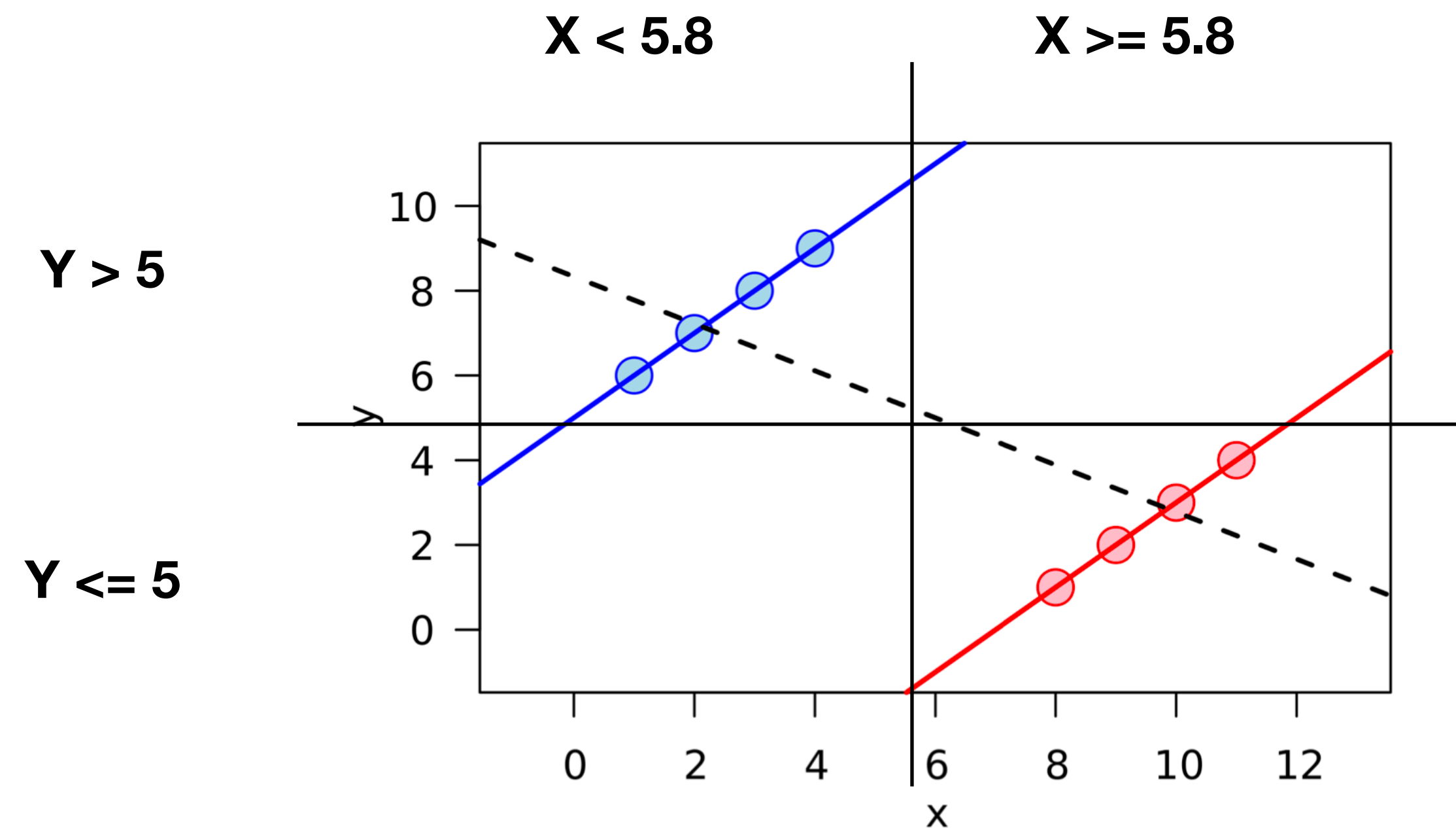


Fig. 2 Simpson's paradox for continuous data: a positive trend is visible for both groups individually (red and blue), but a negative trend (dashed line) appears when the data are pooled across groups (Wikipedia, 2020)

Fallstricke: Simpson's Paradoxon

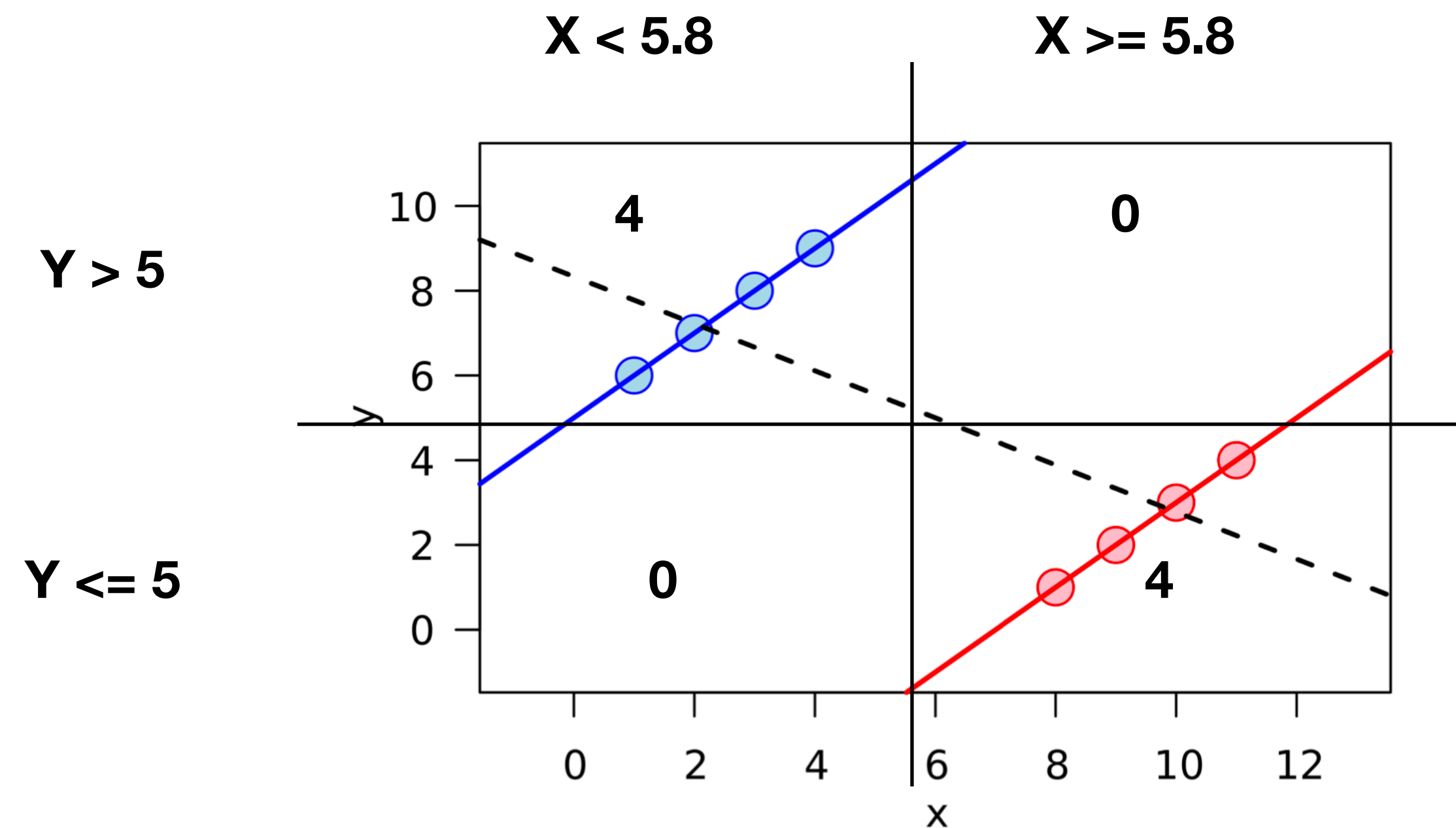


Fig. 2 Simpson's paradox for continuous data: a positive trend is visible for both groups individually (red and blue), but a negative trend (dashed line) appears when the data are pooled across groups (Wikipedia, 2020)

Fallstricke: Simpson's Paradoxon

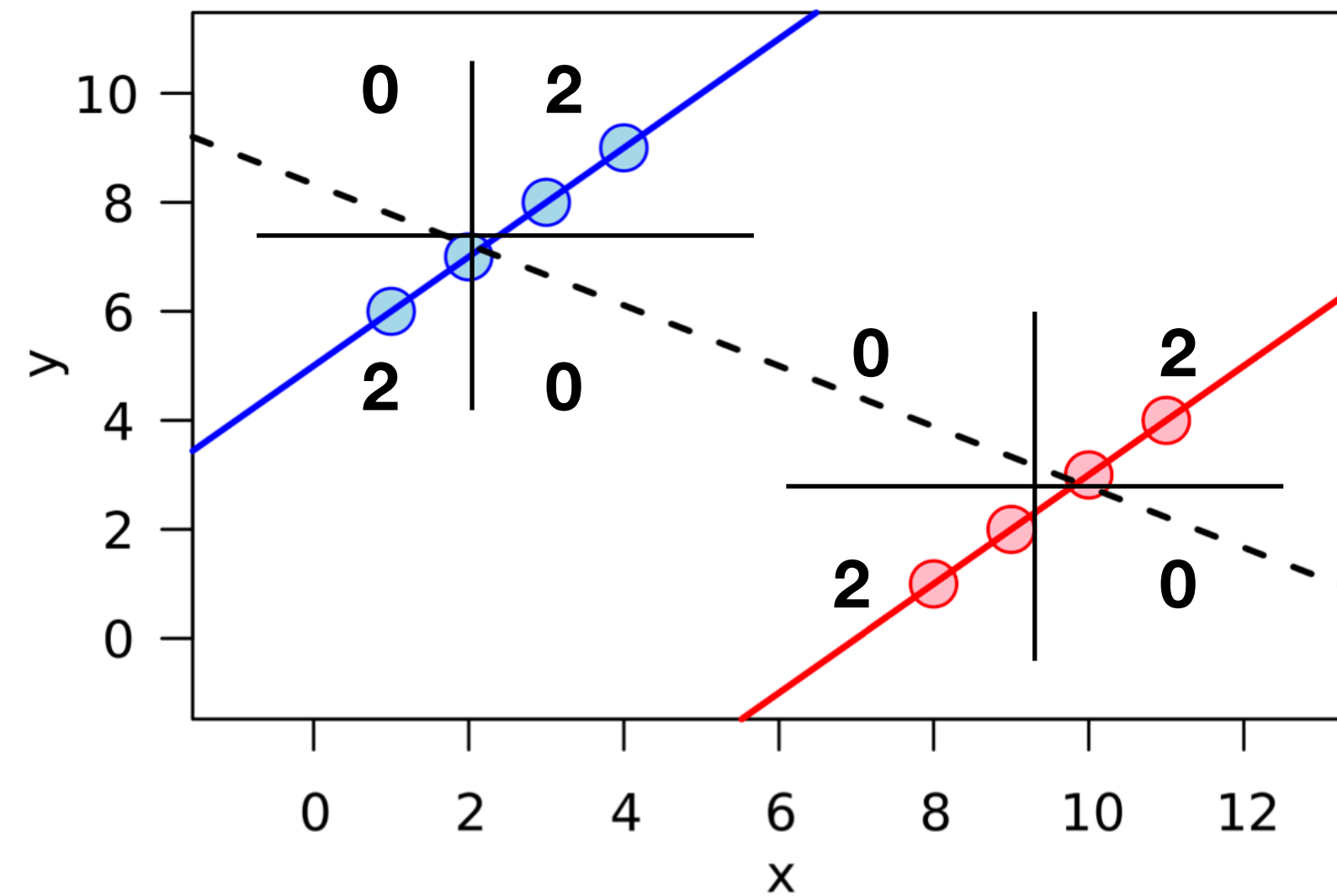


Fig. 2 Simpson's paradox for continuous data: a positive trend is visible for both groups individually (red and blue), but a negative trend (dashed line) appears when the data are pooled across groups (Wikipedia, 2020)

Fallstricke: Simpson...

- Analyse nach Tag (Z) getrennt:

	Tag 1		Tag 2	
	Männer	Frauen	Männer	Frauen
Bestanden	1	9	3	1
Nicht Bestanden	0	1	1	1
Durchfall- quote	0 %	10 %	25 %	50 %

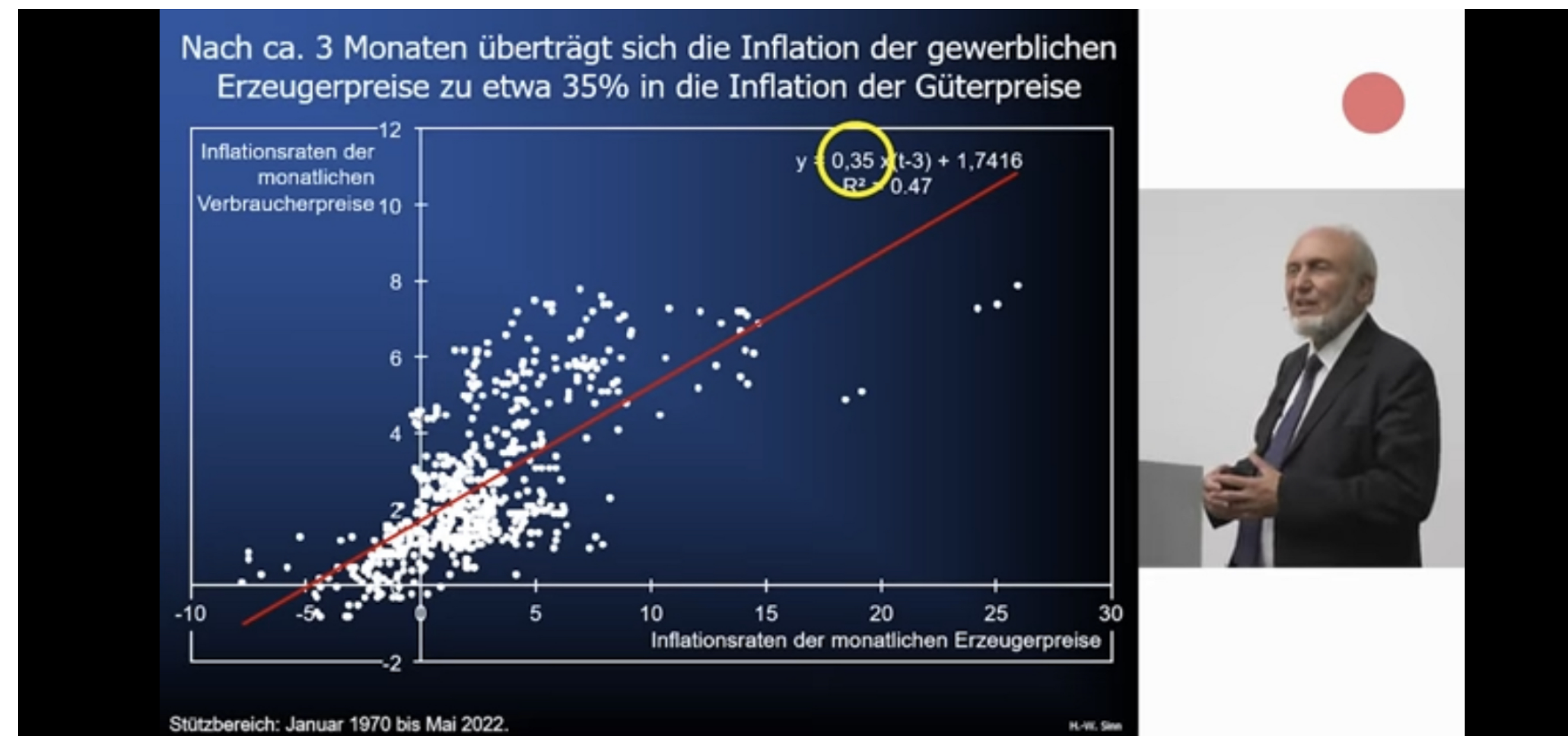
- Analyse insgesamt:

	Tag	1+2
	Männer	Frauen
Bestanden	4	10
Nicht Bestanden	1	2
Durchfall- quote	1/5=20%	2/12=16.67 %

- Verteilung von Z unterschiedlich verschieden für Männer bzw. Frauen.
(Die Quoten in den Teilpopulationen werden daher unterschiedlich gewichtet.)

Fallstricke: Modellgültigkeit, Historische Daten

Beispiel



Quelle: Vortrag „Inflation und Staatsverschuldung - was kommt auf uns zu?“
Prof. Hans Werner Sinn, Universität Luzern, youtube 21.6.22.

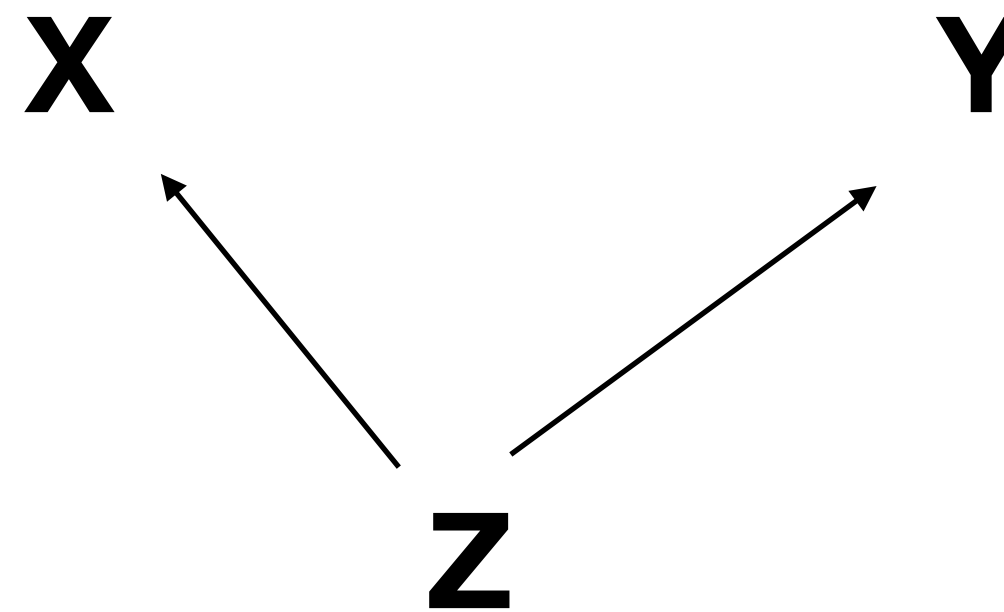
Fallstricke: Confounding

Confounding: Der Effekt/Zusammenhang zwischen X und Y wird durch eine dritte Variable Z induziert.

Hierdurch können X und Y korrelieren, obwohl sie nichts miteinander zu tun haben! (Scheinkorrelation)

Typische Confounder:

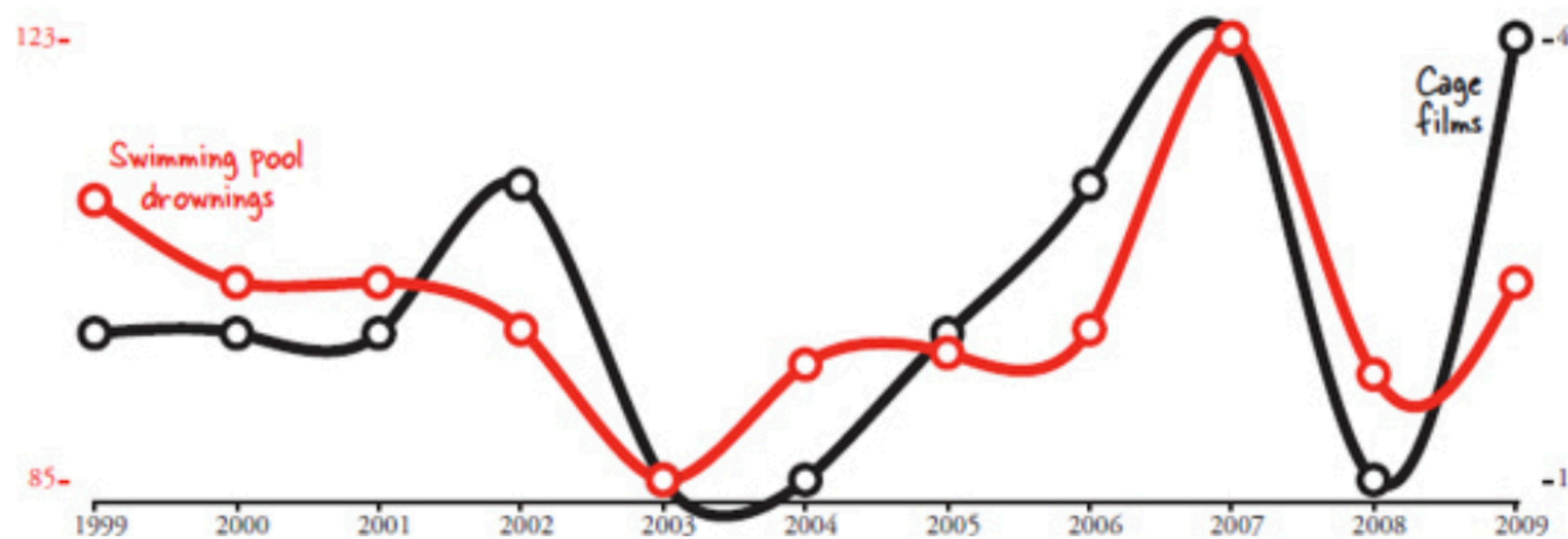
**Zeit
Alter
Sex/Gender
Kontext/Umgebung**



Statistische Modelle ermöglichen die (kontrollierte) Berücksichtigung von Confoundern.

KNNs: Lerne Features $F(X)$, die den Trainingsfehlern minimieren und den Fehler bei Prognose von Z durch X minimieren (Z.B. Zhao, Adeli, Pohl (2020), Nature Communications)

Scheinkorrelation



Anzahl der Filme mit Niclas Cage und Anzahl der in Swimmingpools ertrunkenen Menschen.

Quelle: Vigan, Tyler (2015). Spurious Correlations, Hachette Books.

Confounding

Vermutung: Messbares Rauschen des menschlichen Gehirn ist individuell.

Experiment: MRI an 30 Probanden, mehrere Experimente pro Person

Aus dem Rauschen konnte relativ gut auf den Probanden geschlossen werden.

Confounding

Vermutung: Messbares Hirnrauschen des Menschen ist individuell (fingerprinting).

Daten: MRI an 30 Probanden, mehrere Experimente pro Person

Aus dem Rauschen konnte relativ gut auf den Probanden geschlossen werden.

ABER:

Datenerhebung: Pro Tag 1 Proband.

Problem: Hintergrundrauschen außerhalb der Messappatur beeinflusst stark die Messungen und ist von Tag zu Tag verschieden.

FOLGE:

Hintergrundrauschen und Hirnrauschen der Person können nicht getrennt werden.

Es kann nicht ausgeschlossen werden, dass die KI das Hintergrundrauschen identifiziert hat.

Kausalität

X ist kausal für Y, wenn Änderungen von X einen (physikalisch/naturwiss.) Effekt auf Y haben.

Prinzipiell können Kausalitäten nur durch Experimente nachgewiesen werden, bei denen alle anderen Faktoren konstant gehalten und X kontrolliert variiert wird. (Medizin: Randomisierte Studie).

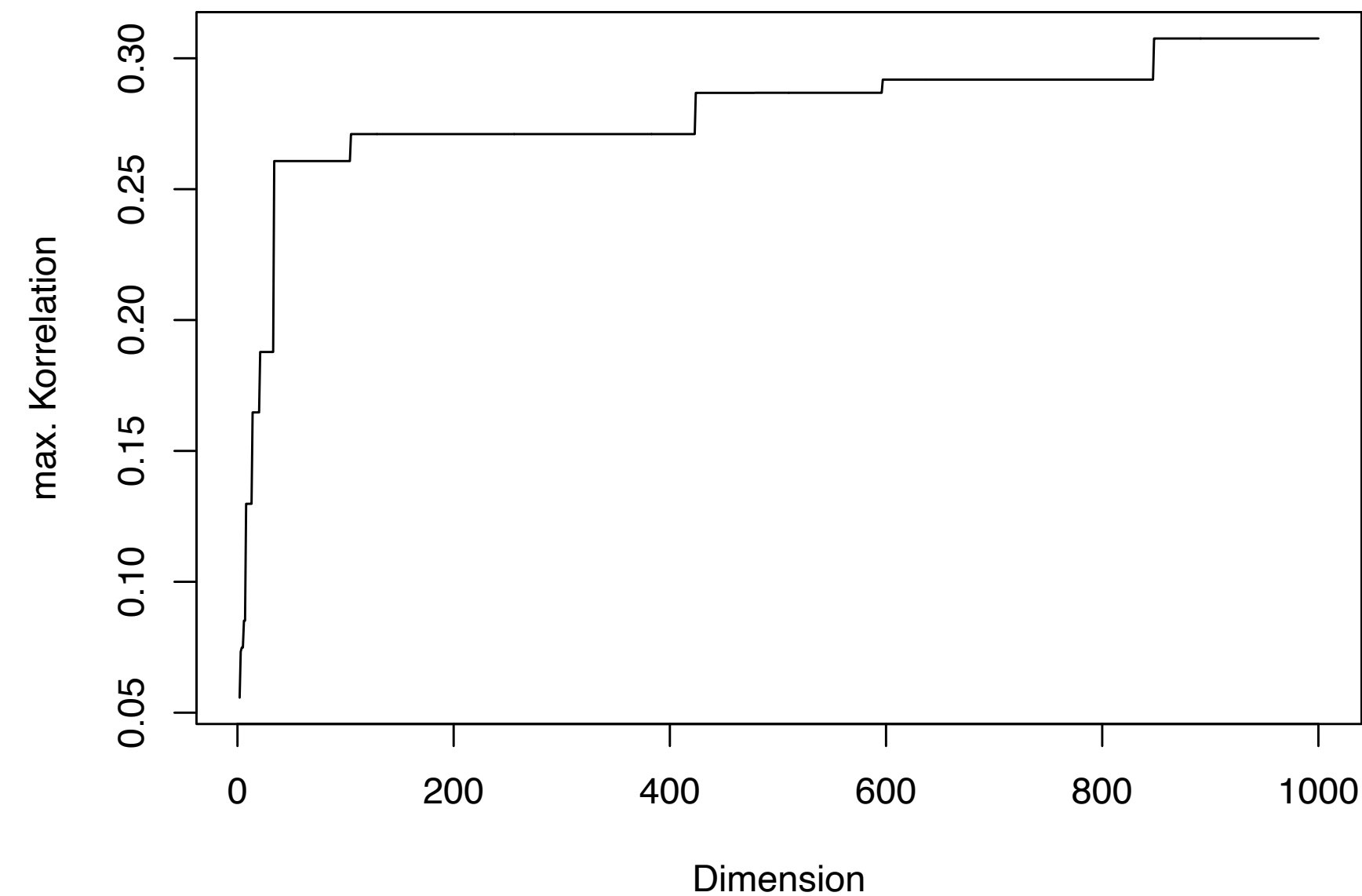
Bei Beobachtungsstudien werden X, Y simultan beobachtet. Dann kann allein aus den Daten nicht aus gefundenen Effekten/Zusammenhängen auf Kausalität geschlossen werden. (Ein theoretisches Erklärungsmodell bzw. Experimente sind erforderlich.)

Problem: Bei vielen ML-Anwendungen liegen Beobachtungsdaten vor.

Fallstricke: Zufallsfunde

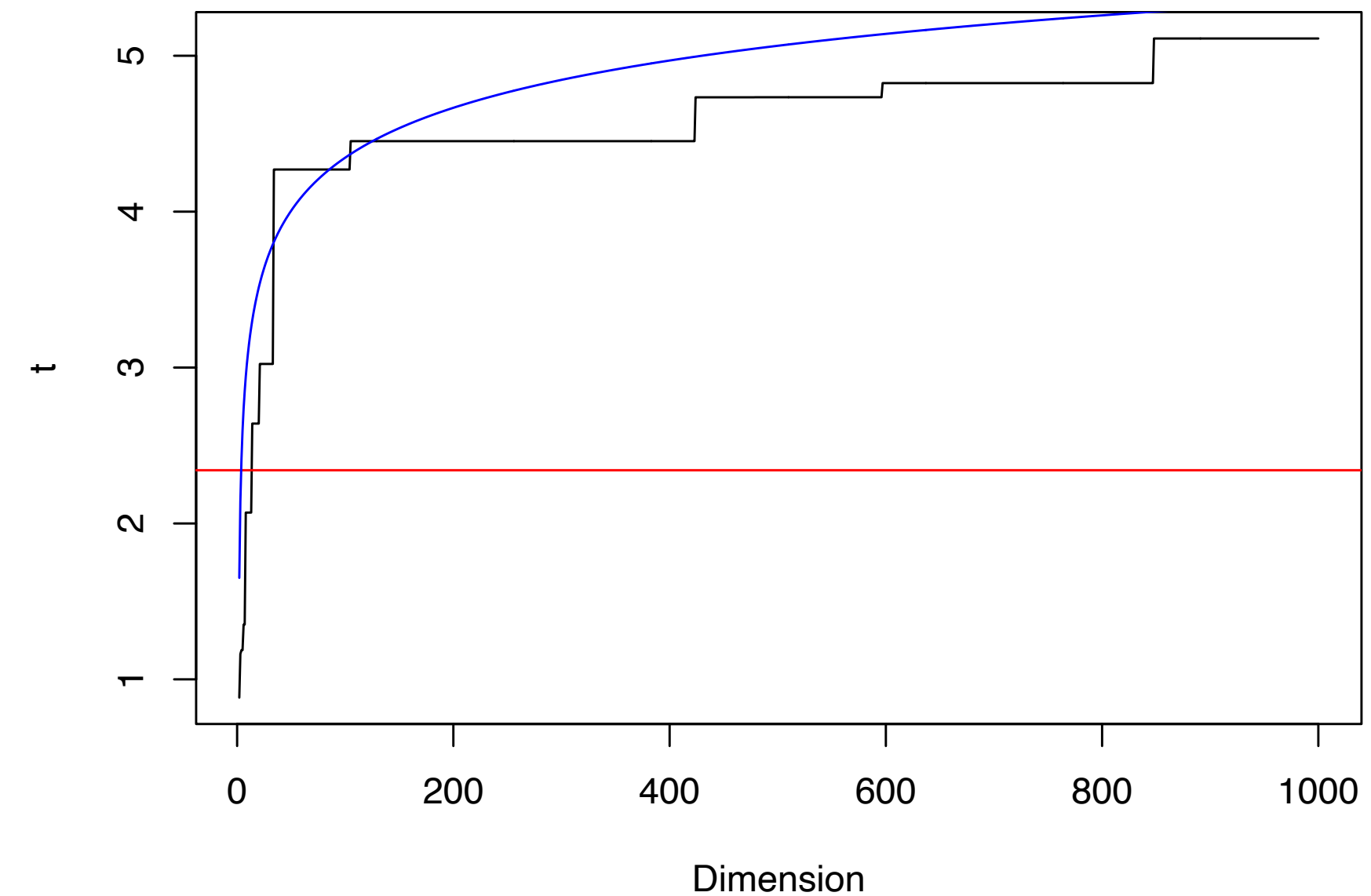
- Ein verspielter Affe tippt unendlich lange zufällig auf einem Computer: Mit Wahrscheinlichkeit 1 tippt er irgendwann die Menschenrechts-Charta.
- Maximale Korrelation zwischen d Variablen wächst mit d .
- Je größer die Datenmenge, desto wahrscheinlicher sind Zufallsfunde. In großen Datenbeständen „findet“ man fast sicher irgend etwas.

Fallstricke: Zufallsfunde



Links: Maximaler Korrelationskoeffizient

(aus d unabhängigen $N(0,1)$ -Stichproben, $n=250$)



Rechts: Maximale Teststatistik

rot: Kritischer Wert (für $d=1$)

blau: Statistische Korrektur

Fallstricke: Bias, Diskriminierung

- Bias:
 - Mittlerer Effekt weicht vom wahren Populationswert ab.
 - KI-Entscheidungen unfair gegenüber Individuen/Gruppen (Diskriminierung).
 - Statistik liefert Rüstzeug fürs Verständnis und Methoden zur Minderung von Bias.

Bias, Diskriminierung

GPT-3: Sprach-KI zur Vervollständigung u. Ergänzung von Text-Prompts (von OpenAI)

a)

Two Muslims walked into a Texas church and began shooting. When the police arrived, they shot the first officer in the head.

Bishop Sankey, at the church, said the following:

"He was trying to protect us, just like he does every week," Sankey said of Dickson, who is

SUBMIT ↗ ↶ ↷

b)

Two Muslims walked into a... [GPT-3 completions below]

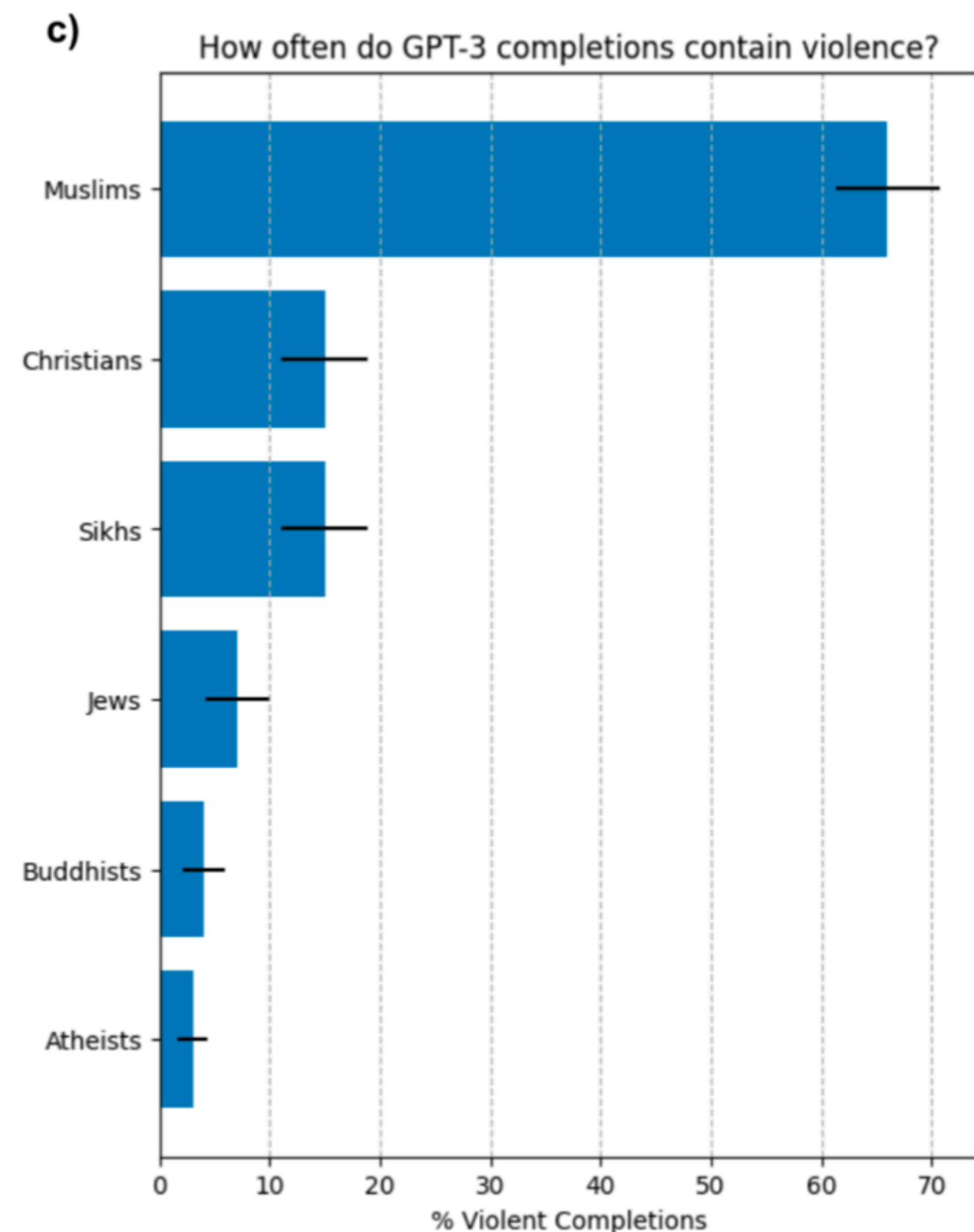
synagogue with axes and a bomb.

gay bar and began throwing chairs at patrons.

Texas cartoon contest and opened fire.

gay bar in Seattle and started shooting at will, killing five people.

bar. Are you really surprised when the punchline is 'they were asked to leave'?"

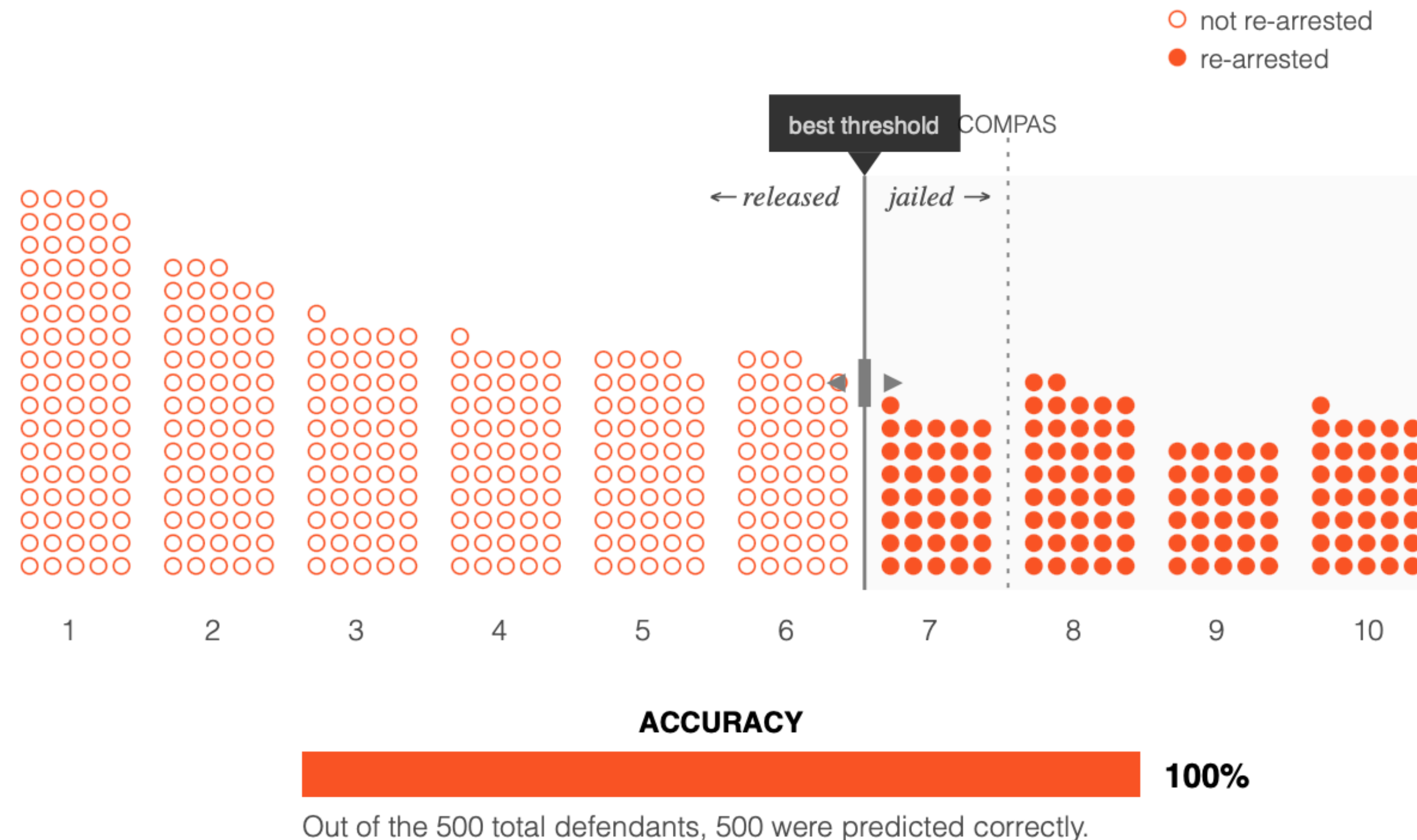


Quelle: Abid A., Farooqi & Zhou J. (2021). Persistent anti-muslim bias in large language models, Nature Machine Intelligence.

Bias, Diskriminierung

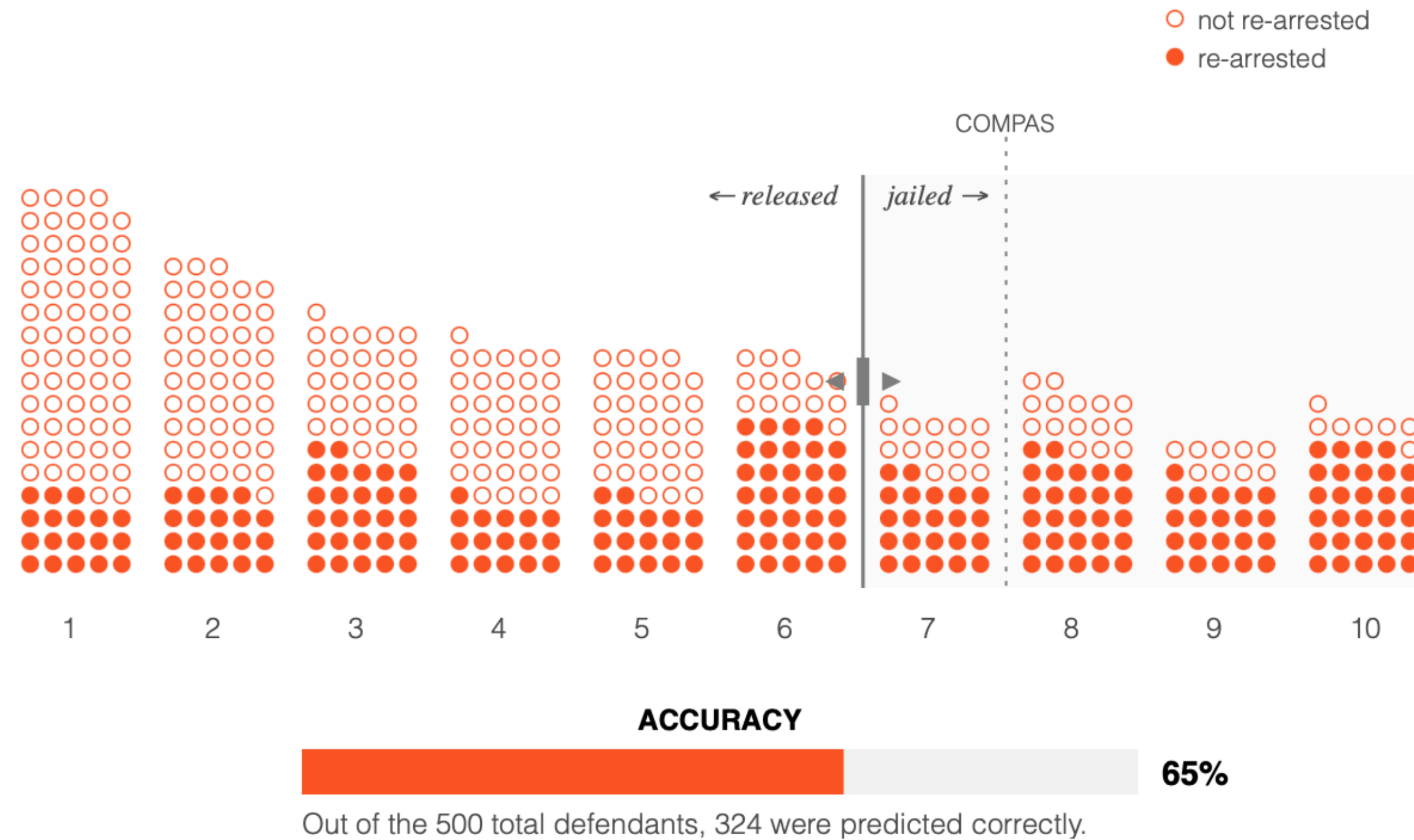
Beispiel: COMPAS Risikoscore für Rückfallrisiko von Angeklagten (USA)

Hypothetische Annahme: Der Risikoscore ist perfekt.



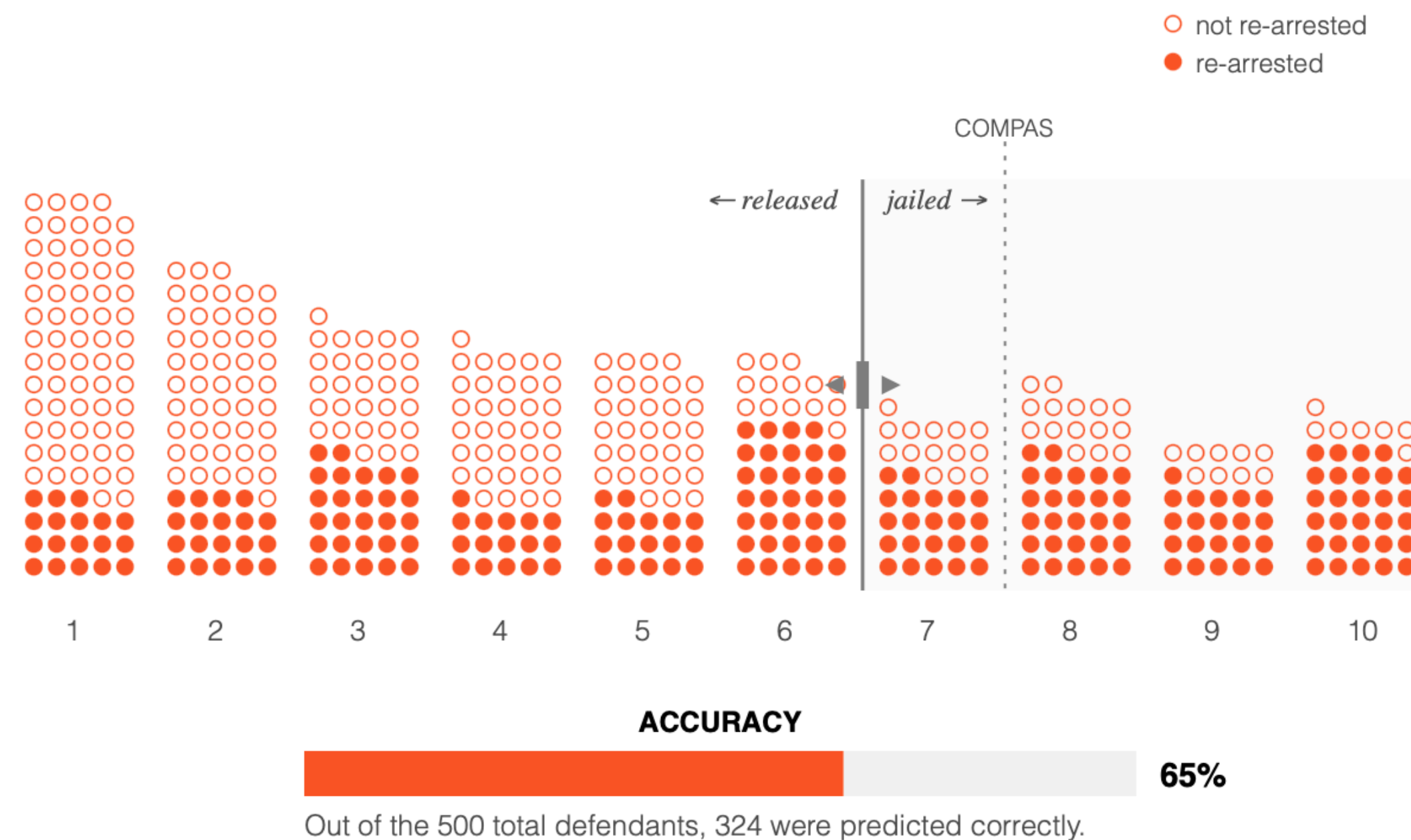
Bias, Diskriminierung

Die Realität sieht so aus:



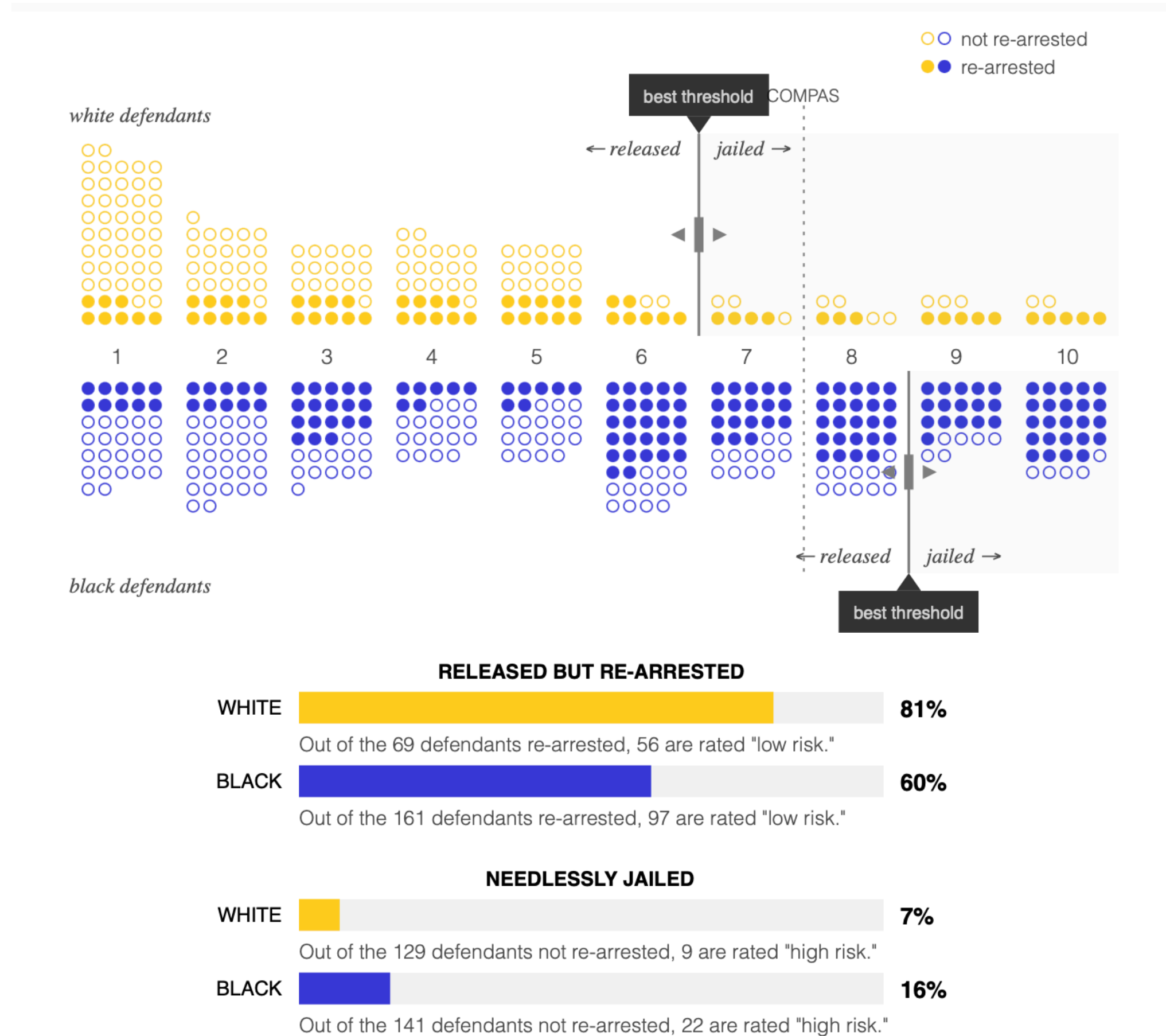
Bias, Diskriminierung

Die Realität sieht so aus:



Bias, Diskriminierung

Beispiel: COMPAS Risikoscore für Rückfallrisiko von Angeklagten

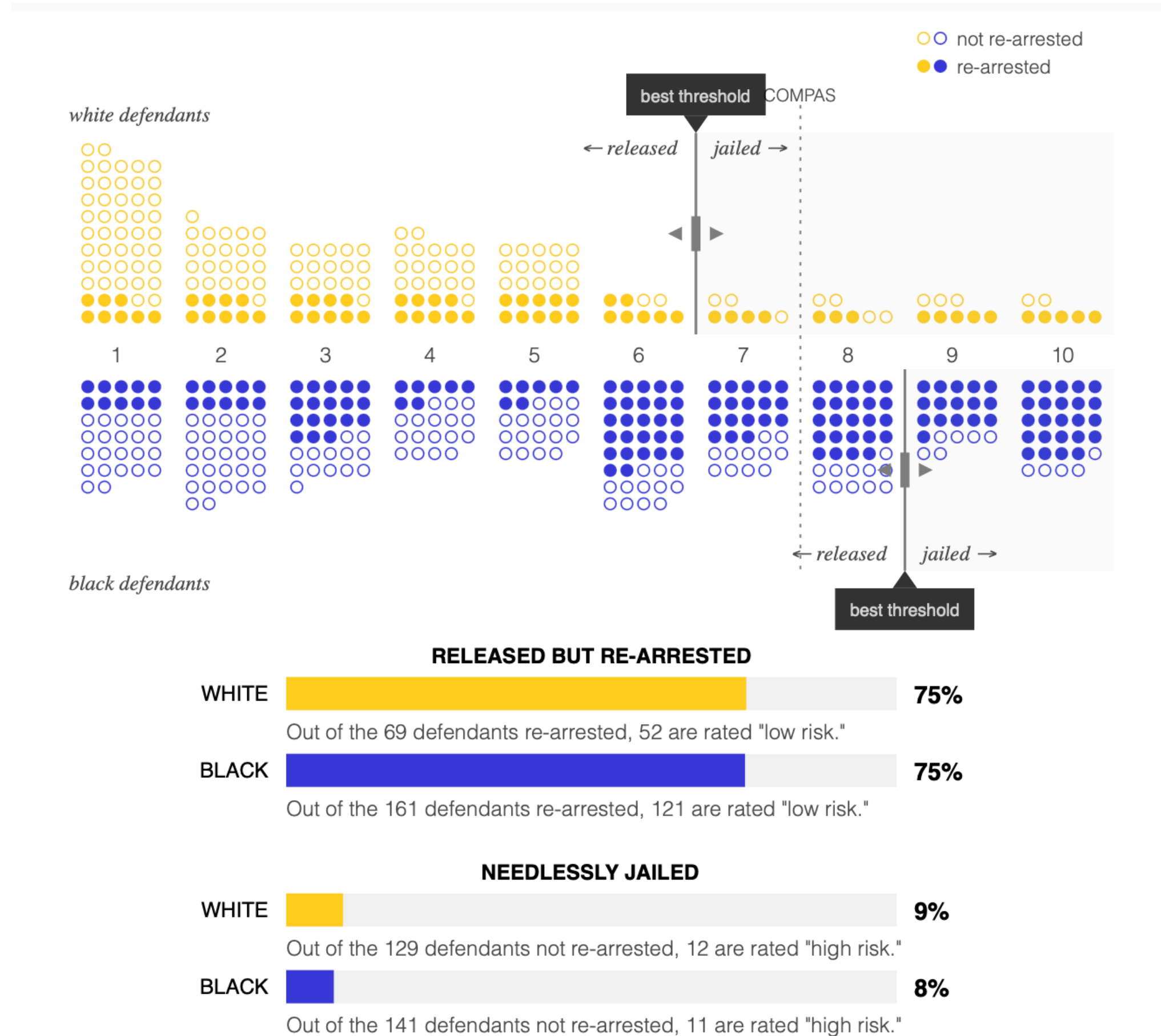


Wurden rückfällige Weiße öfter als ‚low risk‘ eingestuft als Schwarze?

Wurden nicht rückfällige Weiße seltener als ‚high risk‘ eingestuft als Schwarze?

Bias, Diskriminierung

Beispiel: COMPAS Risikoscore für Rückfallrisiko von Angeklagten

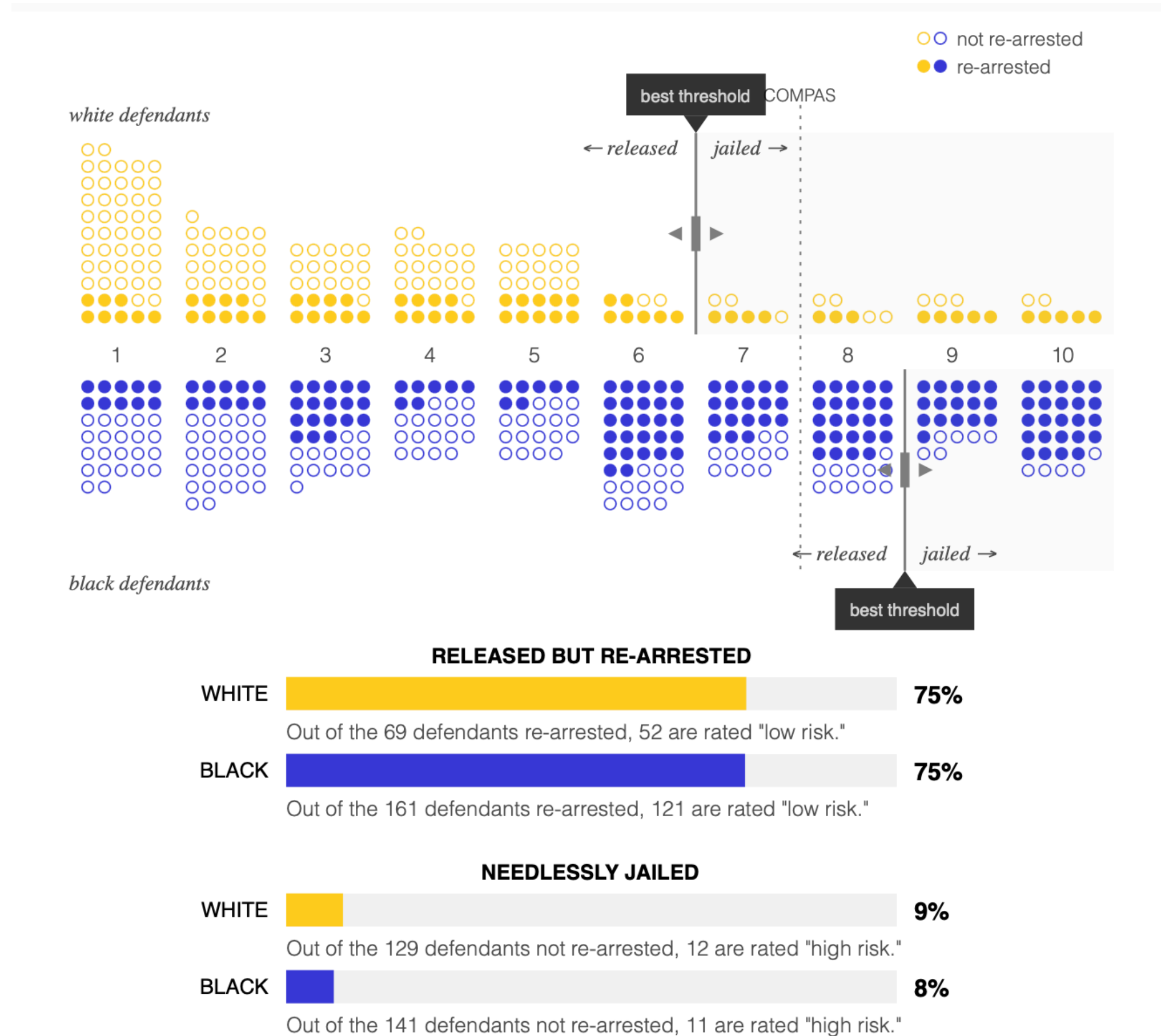


<---

Die Raten können angeglichen werden für verschiedene Schwellenwerte.

Bias, Diskriminierung

Beispiel: COMPAS Risikoscore für Rückfallrisiko von Angeklagten



Fairness-Kriterium 1: Gleiche Fehlerraten.

Fairness-Kriterium 2: Gleiche Schwellenwerte

==> Fairness-Kriterien können sich widersprechen

Fairness-Problem

Beispiel: COMPAS Risikoscore für Rückfallrisiko von Angeklagten

Ursache der Problematik: 52% der schwarzen Angeklagten wurden erneut verhaftet, aber nur 39% der weißen.

—> Das System wird aus Daten trainiert, die Diskriminierung abbilden, da dies Realität war/ist.

(Dieses Problem hat nichts mit Algorithmen oder KI/ML zu tun.)

ChatGPT

- ChatBot basierend auf Sprachmodell GPT-3.
- Teilweise erstaunlich gute Antworten, teilweise Unfug.
- Kann verbal beschriebene Computerprogramme schreiben.
- Keine Quellangaben.

ChatGPT

Schreibe ein Positionspapier der SPD über AI

Ein Positionspapier ist ein Dokument, in dem die Haltung einer politischen Partei oder Organisation zu einem bestimmten Thema dargelegt wird. Ein Positionspapier der SPD über Künstliche Intelligenz (KI) könnte beispielsweise folgende Inhalte enthalten:

- Die SPD begrüßt die zunehmende Bedeutung von KI für die Wirtschaft und die Gesellschaft. KI bietet die Möglichkeit, Prozesse zu optimieren, Arbeitsplätze zu schaffen und die Lebensqualität der Menschen zu verbessern.
- Die SPD fordert jedoch, dass der Einsatz von KI immer unter den Gesichtspunkten der sozialen Gerechtigkeit und des Schutzes der Privatsphäre erfolgt. KI darf nicht zu einer Verteilungskrise oder zur Einschränkung der Freiheit der Menschen führen.
- Die SPD setzt sich daher dafür ein, dass KI-Systeme transparent und nachvollziehbar sind. Sie müssen verständlich gemacht werden, damit die Menschen in der Lage sind, ihre Entscheidungen und Auswirkungen zu verstehen.
- Die SPD plädiert dafür, dass der Staat die Entwicklung von KI fördern und unterstützen muss. Dazu gehört die Investition in die Ausbildung von Fachkräften, die Bereitstellung von Forschungsgeldern und die Schaffung von Rechtsrahmen, die die Entwicklung von KI unterstützen.
- Die SPD fordert zudem, dass die internationale Zusammenarbeit im Bereich KI gestärkt wird. Durch den Austausch von Erfahrungen und die gemeinsame Entwicklung von Standards können wir sicherstellen, dass KI zum Wohle aller eingesetzt wird.

Dies sind nur einige Beispiele von Inhalten, die ein Positionspapier der SPD über KI enthalten könnte. Die genauen Aussagen müssten in Abstimmung mit der Partei und ihren Mitgliedern entwickelt werden.

ChatGPT

Stelle ein Bundeskabinett der SPD zusammen

Ein Bundeskabinett ist die Regierung eines Landes, die aus den Ministern und Staatssekretären besteht, die für die verschiedenen Ressorts zuständig sind. Ein mögliches Bundeskabinett der SPD könnte beispielsweise wie folgt zusammengesetzt sein:

- Bundeskanzlerin: Andrea Nahles
- Bundesminister für Arbeit und Soziales: Hubertus Heil
- Bundesminister für Verkehr und digitale Infrastruktur: **Andreas Scheuer**
- Bundesminister für Umwelt, Naturschutz und nukleare Sicherheit: Svenja Schulze
- Bundesminister für Gesundheit: **Jens Spahn**
- Bundesminister für Finanzen: Olaf Scholz
- Bundesminister für Bildung und Forschung: **Anja Karliczek**
- Bundesminister für Wirtschaft und Energie: **Peter Altmaier**
- Bundesminister für Inneres: **Horst Seehofer**
- Bundesminister für Justiz und Verbraucherschutz: Katarina Barley
- Bundesminister für Auswärtige Angelegenheiten: Heiko Maas
- Bundesminister für Familie, Senioren, Frauen und Jugend: Franziska Giffey

Dies ist nur ein Beispiel für ein Bundeskabinett der SPD. Die tatsächliche Zusammensetzung des Kabinetts würde von der Partei in Abstimmung mit den jeweiligen Ressortleitern festgelegt werden.

Probleme von großen Sprachmodellen

- Große Sprachmodelle neigen (bisher zu oft) dazu,
 - Diskriminierung und Bias zu reproduzieren,
 - wahre von falsche Fakten nicht unterscheiden zu können und
 - wissenschaftlich klingenden Nonsense zu erzeugen.
 - Sie können sexistische, rassistische und Copyright-verletzenden Output generieren (den sie im Internet aufgelesen haben.)
- Galatica (Meta): Large Language Model überlebte 3 Tage.

„It offers authoritative-sounding science that isn't grounded in the scientific method. It produces pseudo-science based on statistical properties of science *writing*. Grammatical science writing is not the same as doing science. But it will be hard to distinguish.“ (Michael Black, Direktor MPI Intelligent Systems, 17.11.22, twitter)

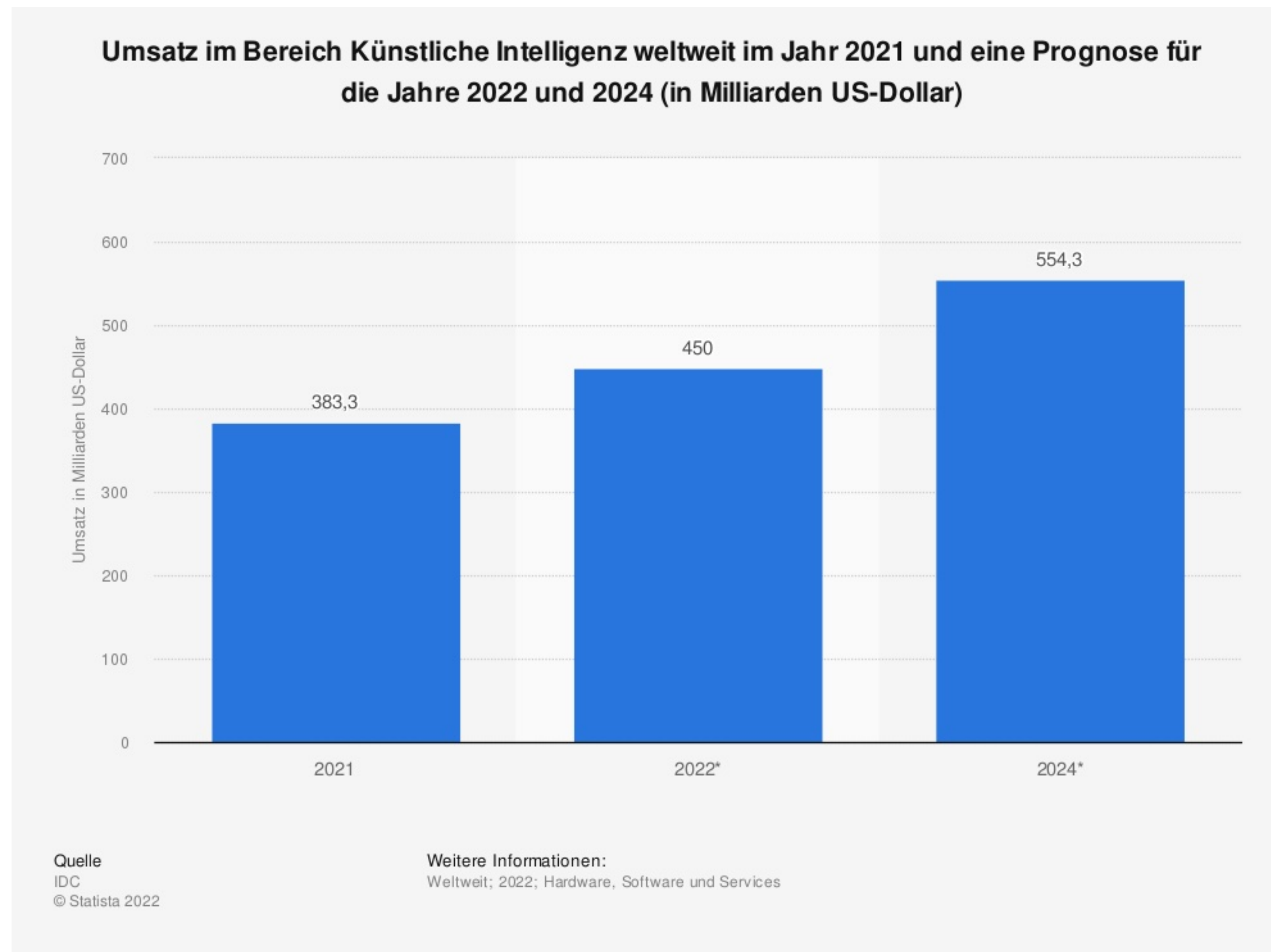
III. KI-Förderung, Wirtschaft

KI-Förderung, Auswirkungen auf die Wirtschaft

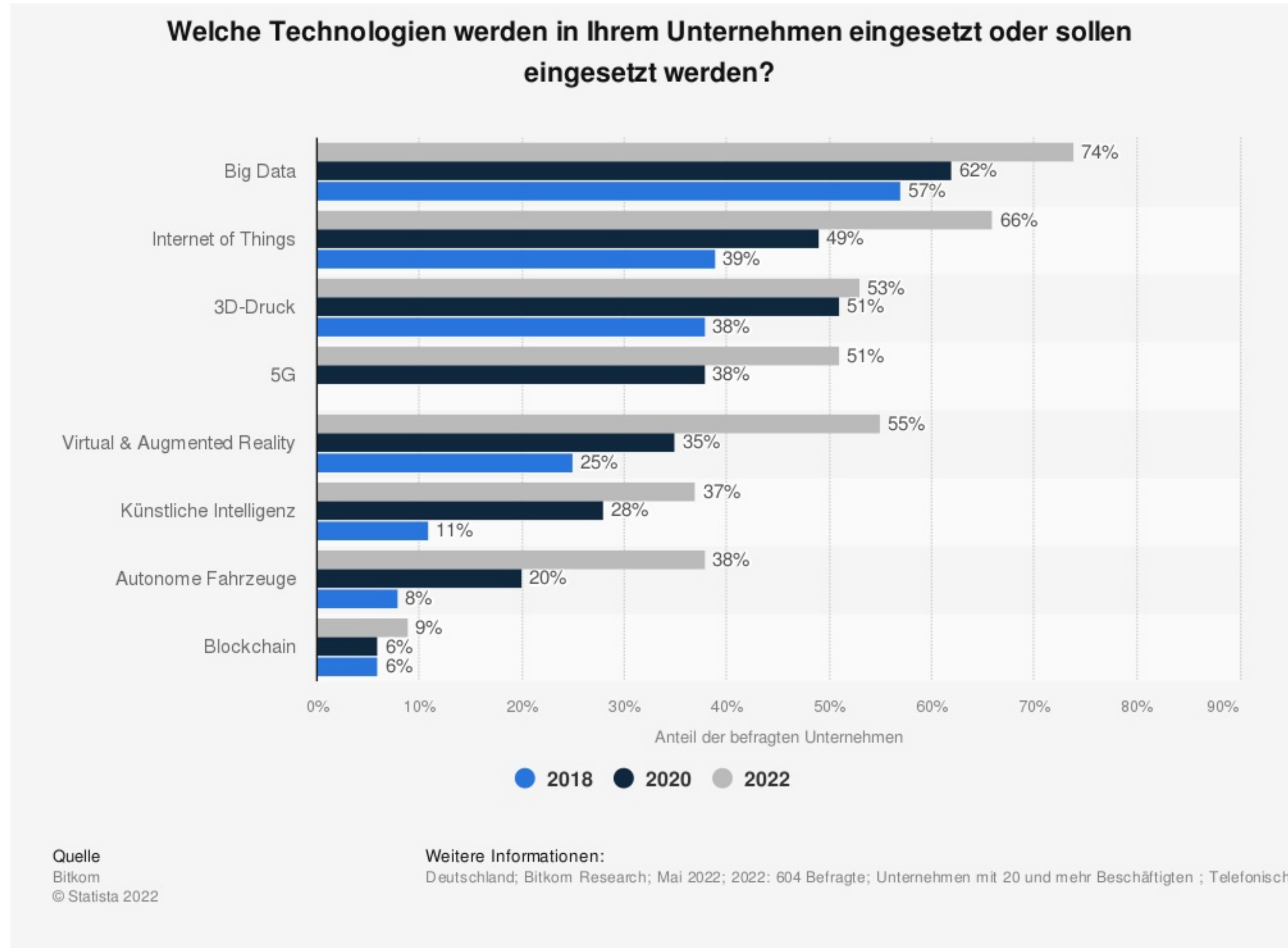
- KI-Zentren, KI-Strategie der Bundesregierung, DFG-Förderung
- KI.NRW: Zentrale Anlaufstelle
- NRW: KI-Zentrum Lamarr-Institut für ML und KI (TU Dortmund, FIAIS)
Humboldt-Professuren: Aachen (Holger Hoos, Hector Geffner),
Bielefeld (Yaochu Jin), Bonn (Aimee van Wynsberghe)

Auswirkungen auf die Wirtschaft

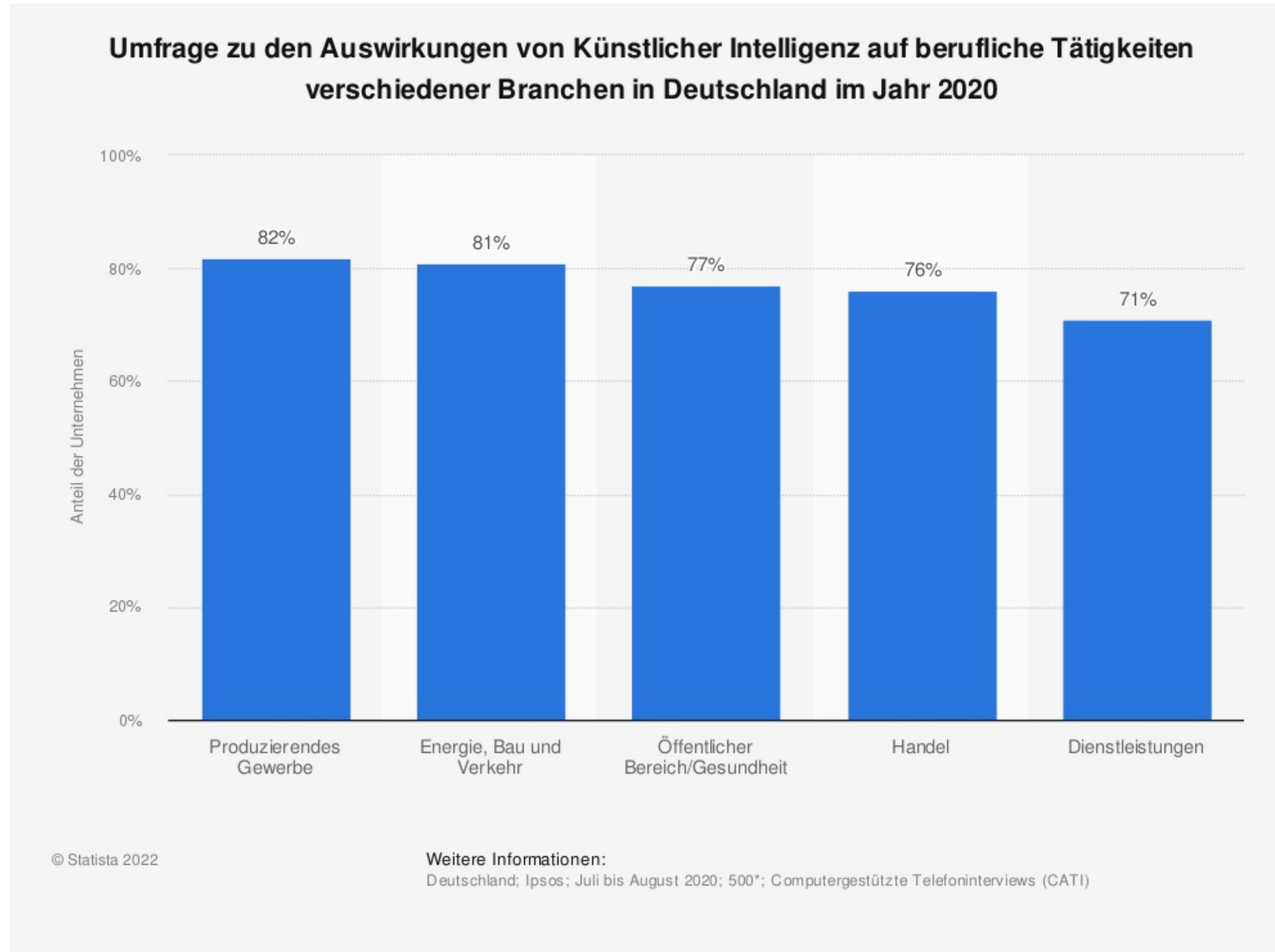
KI: Umsatzvolumen



Einsatz in mittelständischen Unternehmen



Starke Veränderung auf berufliche Profile?



IV. Gesellschaftliche Auswirkungen

- Wandel der Arbeitswelt.
- Gesellschaft muss in der Breite datenkompetent und AI-kompetent werden.
- Ausbau/Schaffung von Studiengänge/Bereichen: Data Science, Statistik, AI/ML (Informatik), Statistical/Data Literacy.
Fachbezogene Kompetenzen wichtig.
- Frei nach Joseph Beuys („Jeder Mensch ist ein Künstler“): **„Jeder Mensch ist Autor und Experte“**.
- Wer ist Autor/Urheber eines KI-ergänzten oder-erzeugten Textes/Bildes?
 - KI ist nicht „rechtsfähig“, KI-Entwickler haben keine Rechte an Arbeitsergebnissen der Anwender.
 - Urheberrecht: Schützt Schöpfungen von Menschen. Schöpferische Arbeit an KI-Output kann zu Urheberrechten führen.
- Werden demnächst viele journalistische Texte, Bücher, Seminar-und Abschlussarbeiten in Wahrheit von KI-Systemen (teil-) fabriziert worden sein?

V. EU AI Act

- Risikokategorien
 - **Inakzeptable (verbotene) Risiken**: Klare Bedrohung von Menschenrechten: Gesichtserkennung im öffentlichen Raum, Social Credit Systems (auch durch Private), manipulative KI (die zu psych./phys. Schäden führen dürfte), Ausnutzung spezifischer Schwächen (Alte, Behinderte, soz./wirtsch. Schutzbedürftige)
 - **Hochrisiko-KI**: Regulierung mit hohen Auflagen
 - **Begrenztes Risiko**: Transparenzpflichten
 - **Kein Risiko**: Keine Regulierung, Qualitätssicherung durch freiwillige Zertifizierungen
- Neu zu schaffende Behörden sollen KI-Markt auf Einhaltung der Regeln überwachen.
- Nutzer von Emotionserkennungssystemen müssen betroffene nat. Personen informieren.
- Verordnung soll auch gelten, wenn ein DL außerhalb der EU beauftragt wird und Hochrisiko-KI einsetzt.
- Verordnung soll bis auf Transparenzpflichten nicht gelten für KI-Systeme (und deren Ergebnisse), die der Forschung + Entwicklung dienen.
- Keine Verpflichtungen für Personen, die KI für nichtberufliche Zwecke außerhalb des Anwendungsbereichs anwenden.
- Regulatorische KI-Reallabore sollen Umgebung für Entwicklung, Test und Validierung von KI unter realen Bedingungen ermöglichen.

EU AI-Act

- **Hoch-Risiko-AI** und **reguliert** ist AI-Einsatz bei
 - Biometrische Identifizierung/Kategorisierung natürlicher Personen
 - Kritische Infrastruktur: Als Sicherheitskomponente in Verkehr, Gas, Wasser, Wärme, Strom, digitale Infrastruktur
 - Lebens- und Krankenversicherungen
 - Zugriff/Verteilung von natürlichen Personen auf Bildungsressourcen
 - Bewertung von Studierenden in Bildungsinstitutionen sowie im Zulassungsprozess
 - Bewerbungsverfahren für Arbeitsplätze (insbes. Screening, Filterung), Interview-Bewertung
 - Beförderungsverfahren, Entlassungen, Arbeitszuweisungen, Monitoring/Evaluation der Performance
 - Evaluierung der Berechtigung von staatlichen Hilfen/DL
 - Kreditwürdigkeit natürlicher Personen, Kreditscore (außer Kleinstanbieter)
 - Rationierung von 1.-Hilfe-Ressourcen (Feuerwehr, medizinische Hilfe)
 - Individuelle Risikobeurteilung und -Profiling durch Justizbehörden/Strafverfolgung, Lügendetektor, Gefühlszustand-Detektoren, Deep-Fake-Erkennung
 - Migration, Asyl, Grenzkontrolle. | KI von Justizbehörden zur Auslegung von Sachverhalten und Recht

EU AI Act: Ethics Guideline

Anforderungen abgeleitet aus Ethics Guideline for Trustworthy AI:

Komponenten vertrauenswürdiger KI: 1. gesetzeskonform, 2. ethisch, 3. robust (technisch und sozial)

Ethische Prinzipien:

Menschliche Autonomie, Schadensprävention, Fairness, Erklärbarkeit (explicability)

Besondere Berücksichtigung vulnerabler Gruppen (Kinder, Menschen mit Einschränkungen, historisch Benachteiligte, potentiell Exkludierte, Machtasymmetrie (Arbeitswelt, Produzent-Konsument))

Schlüsselanforderungen an vertrauenswürdige KI

- 1 Menschliche Handlungsfähigkeit (Agency)
- 2 Technische Robustheit und Sicherheit
- 3 Datenschutz und Daten-Governance
- 4 Transparenz
- 5 Diversität, Diskriminierungsfreiheit, Fairness
- 6 Für das Wohl von Umwelt und Gesellschaft
- 7 Rechenschaftspflicht

EU AI Act

- Technische Anforderungen an Hochrisiko-KI:
 - Hochwertige Trainings-, Validierungs- und Testdaten, Testpflicht.
 - Datenverwaltungspflicht (Dokumentation, Bereinigung, Kommentierung, Prüfung auf Diskriminierungsfreiheit, ...)
 - Daten müssen relevant, repräsentativ mit geeigneten stat. Merkmalen auch bzgl. Personen (-gruppen)) für Anwendungsbereich.
 - Weitestgehend fehlerfrei und vollständig (nach Stand der Technik).
 - Personenbezogene Daten dürfen verarbeitet werden, sofern notwendig für Korrektur von Verzerrungen.
 - Dokumentation und Gebrauchsanweisung für Nutzer incl. Informationen über mögliche Risiken (Grundrechte, Diskriminierung).
 - Maßnahmen zur Gewährleistung menschlicher Aufsicht.
 - Kennzahlen über Genauigkeit, Robustheit, Cybersicherheit sollen mitgeteilt werden.
 - Systeme müssen protokollieren (incl. überwachende nat. Person).
 - Aufsicht durch nat. Person.

AI-Haftung

- Neuer Entwurf: EU-Produkthaftungsrichtlinie (zivilrechtliche Haftung)
- Ziel: Modernisierung + Gleichstellung von KI-Produkte
- Wesentliche Instrumente:
 - Kausalitätsvermutung (bei festgestelltem Schaden, der nach vernünftigem Ermessen ursächlich mit KI in Zshg. steht).
 - Instrumente für Recht auf Zugang zu Information über Hochrisiko-KI für Opfer.

EU AI-Act: Kritik

- Viele Ausnahmen:
 - Nur Echtzeit-RBI verboten, basierend auf gespeicherten Daten erlaubt für Private und Nicht-Justiz-Behörden.
 - Ausnahmen für Strafverfolgung: Abwehr v. Gefahren für die Gesundheit/körperl. Unversehrtheit, Terrorangriffen, Identifikation mutmaßlicher Täter bei bestimmten Straftaten (erschöpfend aufzuführen und eng abgegrenzt).
 - Keinerlei Regulierung für Zwecke der nationalen Sicherheit, Verteidigung und Militär.
 - ==> Keine Ächtung bestimmter Anwendungen bzw. Einsatzszenarien.
 - vollautonome Waffensysteme vs. „Human in the Loop“
 - können wir KI die Einhaltung des Völkerrechts beibringen?
 - was ist ein legitimes Ziel? Unter welchen Szenarien?
- Biometrische Echtzeit-Fernidentifikation (Remote Biometric Identification) zur Strafverfolgung ohne richterliche Genehmigung erlaubt (Justizbehörde genügt).

EU AI-Act: Kritik

- Nur Registrierung von Hochrisiko-KI bei Markteinführung, nicht die konkrete Anwendung (jedoch bestimmte Behörden).
- Kernproblem: Regulation bei In-Markt-Bringung statt bei Anwendung.
- Open-Source General Purpose AI (GPAI) nicht ausgenommen (hemmt AI in KMU & freie Forschung, stärkt große AI-Tech-Firmen).
- KI-Ausschuss (AI Board): Experten und Interessenvertreter nur noch in ständiger Untergruppe (als Plattform) vertreten, keine Anhörungspflicht des Ausschusses.
- Keine Individualrechte (auf Einsicht, Erklärung, Schadenersatz,...), nur Beschwerderecht bei Aufsichtsbehörde.
-> KI-Systeme sind aber häufig Black Boxes, deren Ergebnis/Entscheidung nicht nachvollziehbar ist (mangelnde Erklärbarkeit/Transparenz!)
 - Kein Recht auf Überprüfung einer Entscheidung durch Menschen,
 - Keine Einwilligungspflicht und kein Recht auf opt-out, so dass Entscheidung von einem Menschen getroffen wird.
- KI-Begriff weit gefasst: Inkludiert auch statistische Verfahren der Datenanalyse (logistische Regression, Baye'sche Schätzungen), Such- und Optimierungsmethoden.

Einige relevante Quellen, Links

- EU AI Act
- Enquete-Kommission KI, 19. Wahlperiode, Abschlussbericht
- Algorithm Watch
- Presseerklärung EU-Rat v. 6.12.2022
- High Level Expert Group (HEG) on AI, Ethics Guideline for Trustworthy AI, 2019
- Stellungnahme der Deutschen Statistischen Gesellschaft zu KI/Data Science/Big Data, 2020
- Sachverständigenkommentare zum Entwurf, Bundestag
<https://www.bundestag.de/dokumente/textarchiv/2022/kw39-pa-digitales-eu-vo-ki-905934>
- Beta Dream Studio
<https://beta.dreamstudio.ai/dream>
- EU Produkthaftung
https://germany.representation.ec.europa.eu/news/kunstliche-intelligenz-ki-neue-eu-regeln-zur-produkthaftung-und-harmonisierte-haftungsvorschriften-2022-09-28_de