

Ein Algorithmus hat kein Taktgefühl

Prof. Dr. KA Zweig,
TU Kaiserslautern



The background of the image is Raphael's famous fresco, 'The School of Athens'. It depicts a group of ancient Greek philosophers in a grand, classical architectural setting with arches and statues. The figures are engaged in various activities of study and debate, such as Plato pointing upwards and Aristotle gesturing downwards. The scene is filled with intellectual activity, representing the foundations of Western thought.

Data Science + Ethik?

Netzwerkanalyse





Zentralitätsmaße

Betweenness Centrality

$$betw(v) = \sum_{s,t \neq v} \frac{\delta_{st}(v)}{\delta_{st}}$$

Die *betweenness centrality* wird wie folgt definiert:

δ_{st} steht für die Anzahl kürzester Pfade zwischen s und t .

$\delta_{st}(v)$ steht für die Anzahl kürzester Pfade zwischen s und t , über v .

A person wearing a black balaclava and a dark jacket is crouching against a dark wooden wall. They are holding a rifle with a scope. The image is in grayscale, with red text overlaid.

Und nun zu etwas anderem!

Wie findet man eigentlich terroristische Kuriere?

A close-up photograph of a hand holding a smartphone. The phone's screen is illuminated with a solid blue color. The background is dark and out of focus. Overlaid on the lower half of the image is white text in a sans-serif font.

Gegeben: Handydaten von
55 Millionen Einwohnern aus
Pakistan und Afghanistan



Match made in heaven?



Oder doch
eher
gefährlicher
Unsinn?

Versteckte Annahmen der Betweennesszentralität

$$C_B(v) = \sum_{s,t \neq v} \frac{\delta_{s,t}(v)}{\delta_{s,t}}$$

Uoh...

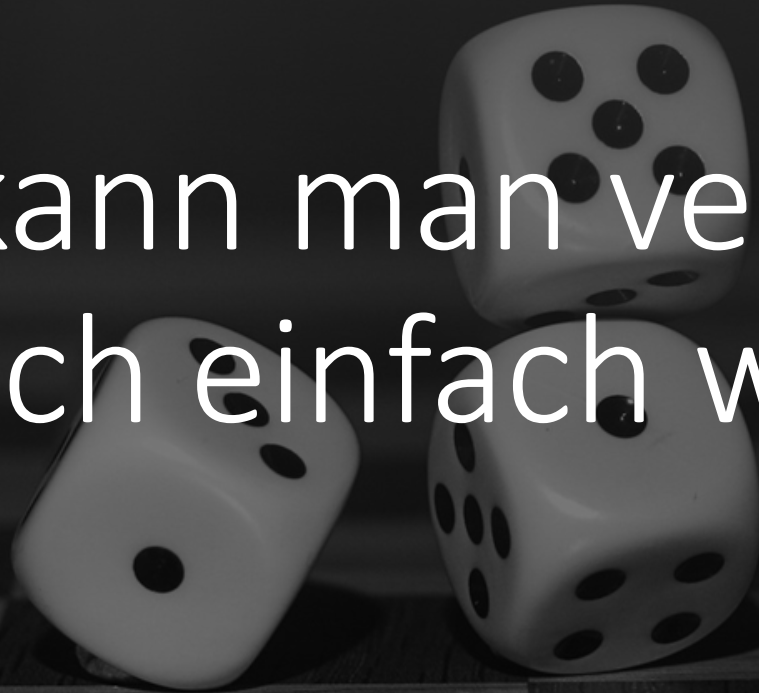
Nur kürzeste Pfade

Okay, das ist eine
Approximation,
richtig?

Dorn et al., 2012
Zweig, 2016

Alle Knotenpaare wollen mit
der selben Frequenz/Intensität
kommunizieren.

Da kann man vermutlich
auch einfach würfeln





Wer ist für die
Fehlanwendung
verantwortlich?

Data Scientist?
Domänenexperte?



Das läuft dann eher so

Minimalforderung: Kommunikati

Vertex and edge betweenness centrality

Description

The vertex and edge betweenness centrality measures the number of shortest paths through a vertex or an edge.

The

Was passiert bei nicht
zusammenhängenden Graphen? Wann
kann Betweenness Centrality inhaltlich
sinnvoll verwendet werden?

be g_{ij} is the number of shortest paths between vertices i and j while g_{ij} is the number of those shortest paths which pass through vertex v .

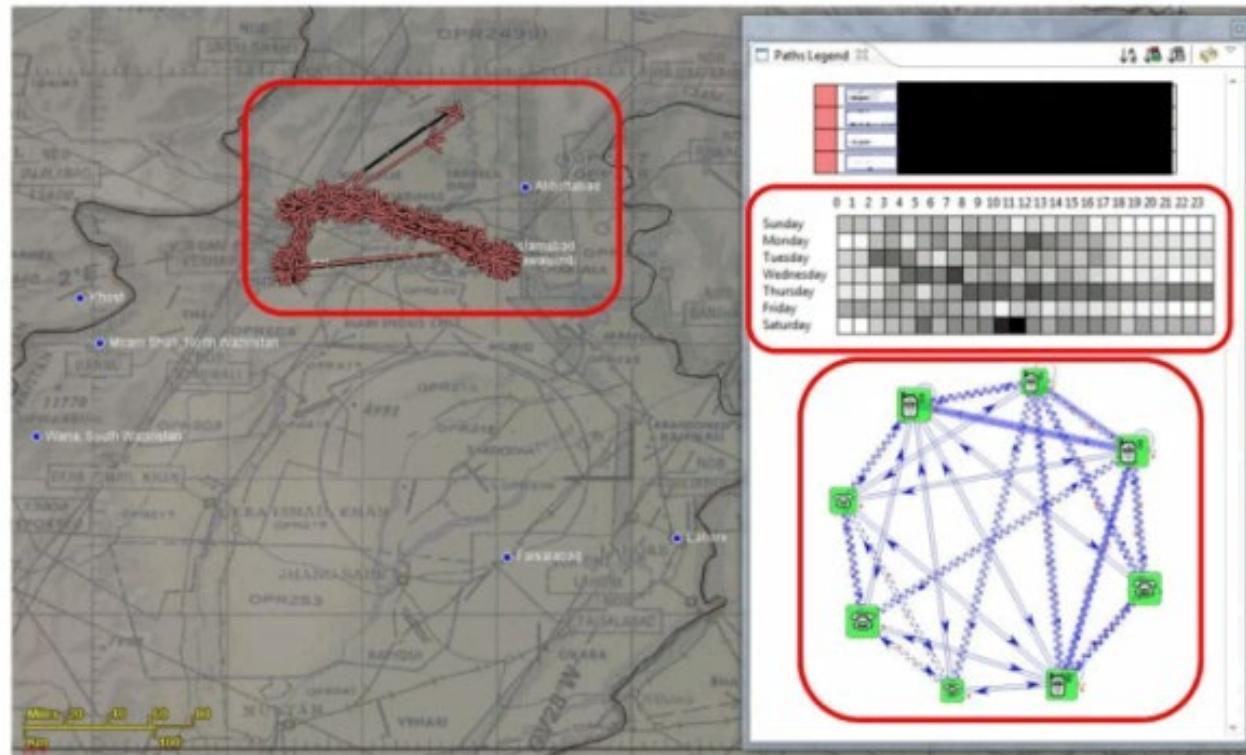
edge_betweenness calculates edge betweenness.

Here g_{ij} is the number of shortest paths between vertices i and j while g_{ivj} is the number of those shortest paths which pass through vertex v .

Fangen von Terroristen mit Netzwerkanalysen

TOP SECRET//COMINT//REL TO USA, FVEY

From GSM metadata, we can measure aspects of each selector's **pattern-of-life**, **social network**, and **travel behavior**



TOP SECRET//COMINT//REL TO USA, FVEY

Terroristen Identifikation SKYNET

TOP SECRET//COMINT//REL TO USA, FVEY

We've been experimenting with several error metrics on both small and large test sets

Training Data	Classifier	Features	100k Test Selectors		55M Test Selectors	
			False Alarm Rate at 50% Miss Rate	Mean Reciprocal Rank	Tasked Selectors in Top 500	Tasked Selectors in Top 100
None	Random	None	50%	1/23k (simulated)	0.64 (active/Pak)	0.13 (active/Pak)
Known Couriers	Centroid	All	20%	1/18k		
		Outgoing	43%	1/27k		
+ Anchory Selectors	Random Forest		0.18%	1/9.9	5	1
			0.008%	1/14	21	6

Random Forest trained on Known Couriers + Anchory Selectors:

- 0.008% false alarm rate at 50% miss rate
- 46x improvement over random performance when evaluating its tasked precision at 100

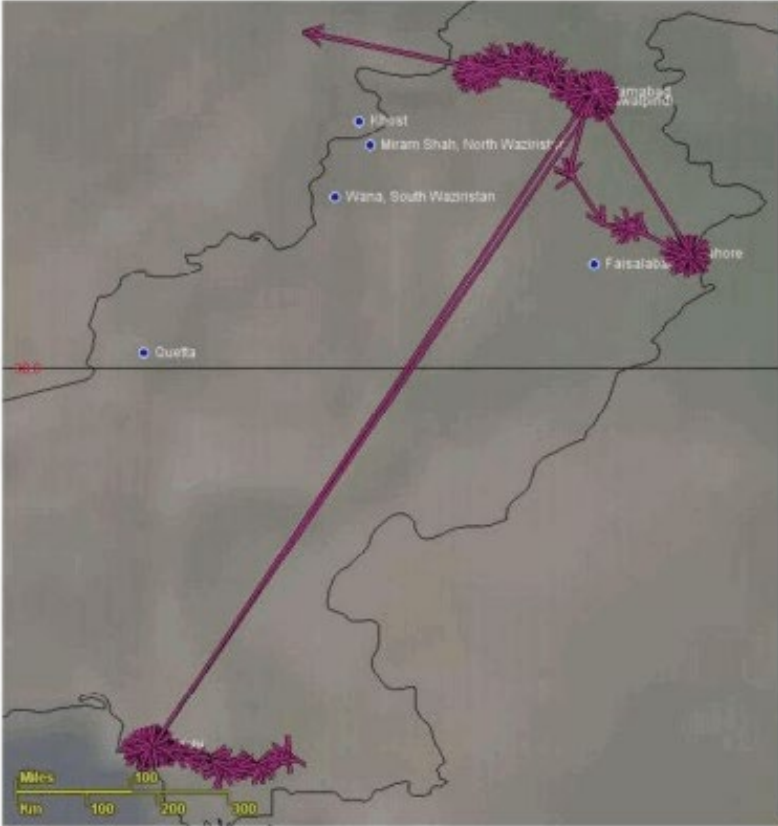
Windo
Wechseln
aktivieren

TOP SECRET//COMINT//REL TO USA, FVEY


Top-"Terroristen-Kurier" ist...

TOP SECRET//COMINT//REL TO USA, FVEY


The highest scoring selector that traveled to Peshawar and Lahore is **PROB AHMED Z Aidan**



The map displays travel routes in purple with starburst markers at key locations. The routes connect Peshawar to Lahore, and then from Lahore to various locations in North and South Waziristan, including Miram Shah, Wana, and Faisalabad. A scale bar at the bottom left indicates distances in miles and kilometers.



PROB AHMED MUWAFAK ZAIDAN



TIDE Person Number: [REDACTED]

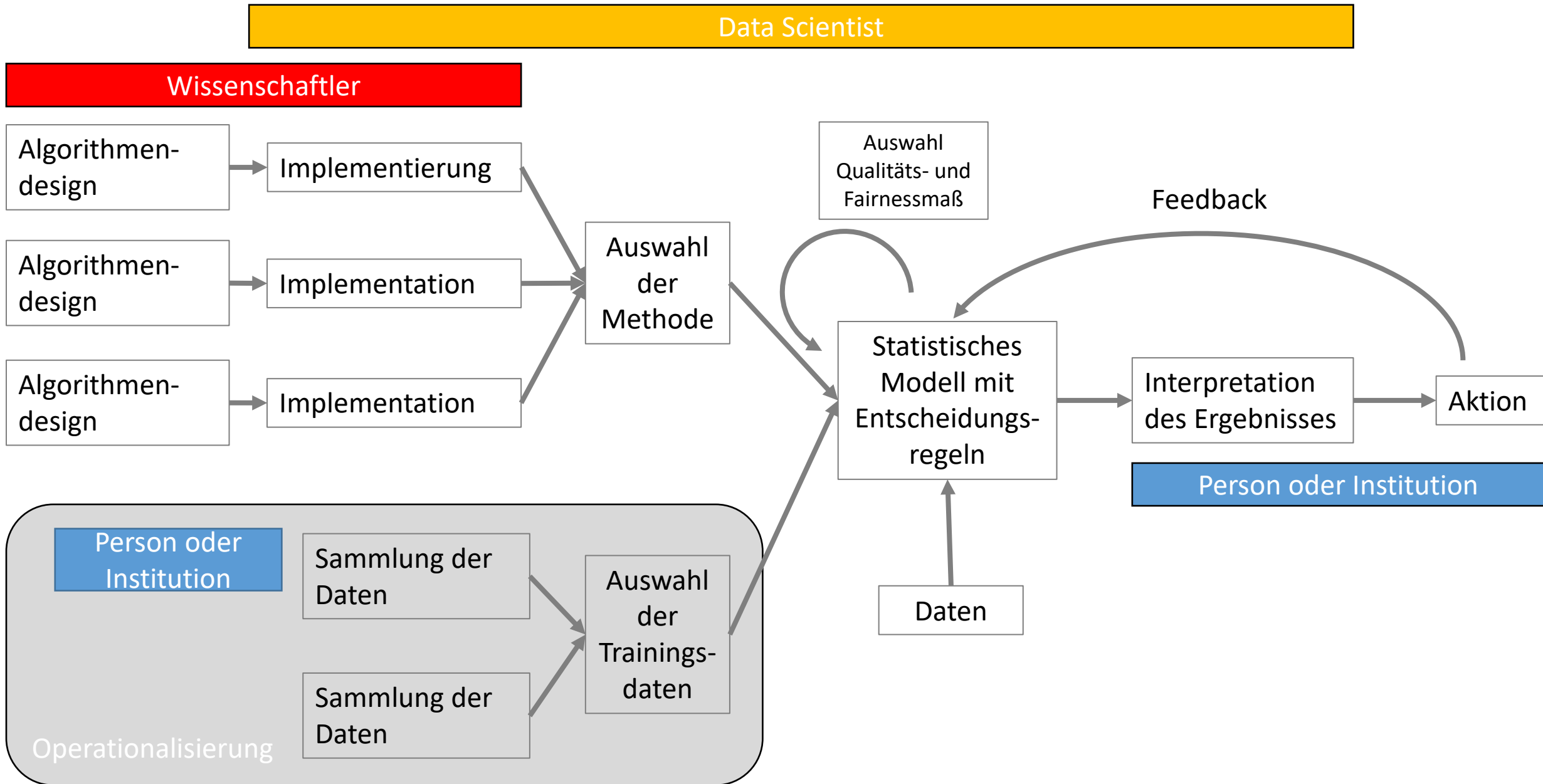
- ~~MEMBER OF AL QAEDA~~
- ~~MEMBER OF HAKIM~~
- ~~MEMBER OF~~
- WORKS FOR AL JAZEERA

Windows
Wechseln

Wer ist verantwortlich für
Fehlanwendung?

Auf dem Holzweg

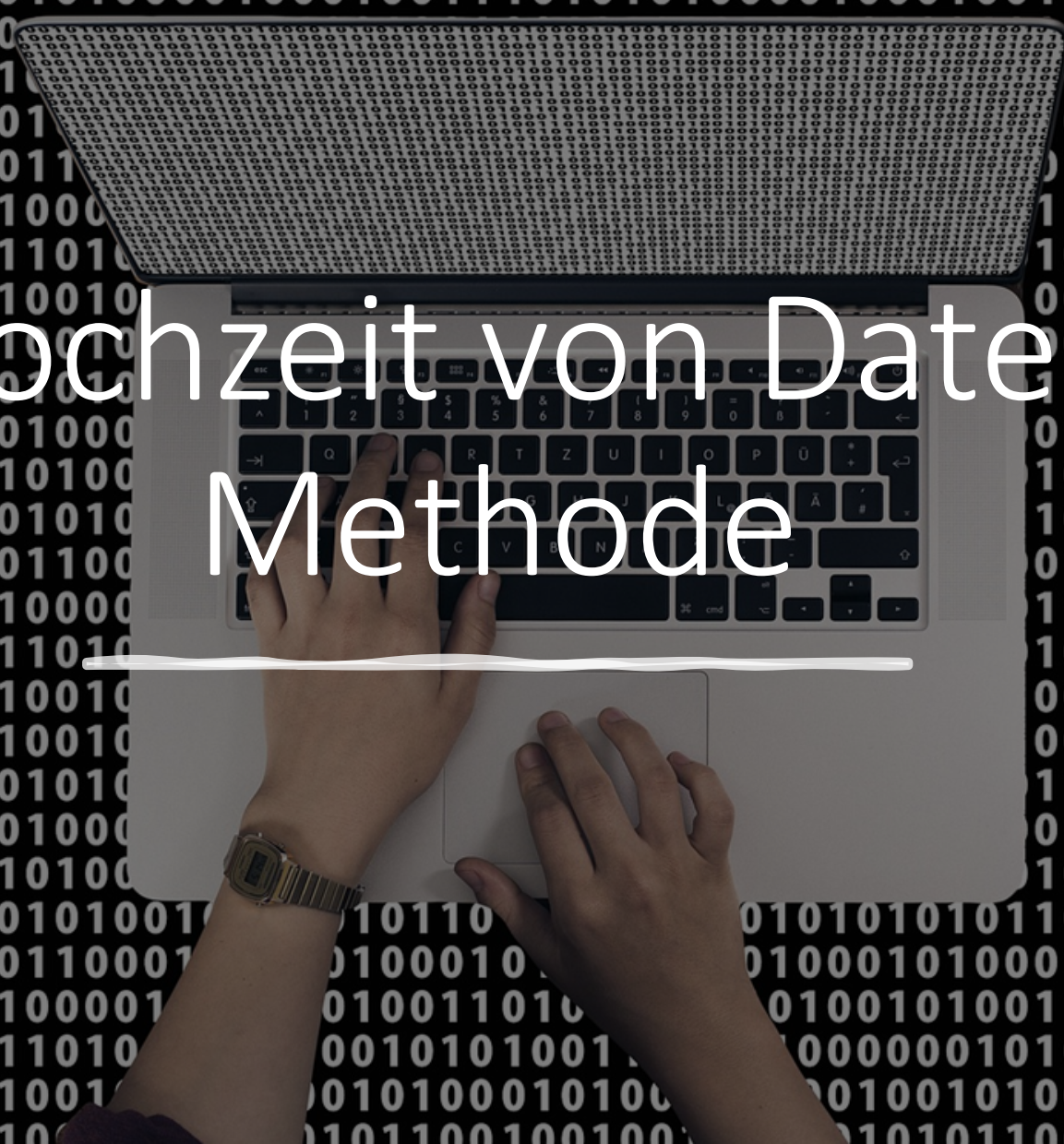




Statistische Modelle sind
emergente Phänomene



1) Hochzeit von Daten & Methode





2) Emergenz

Stau, Massenpanik, Hypes, Moden, Epidemien, Schwarmverhalten....

A close-up photograph of two hands being washed with soap. The hands are covered in thick, white, bubbly suds. The lighting is dramatic, highlighting the texture of the skin and the individual bubbles of the soap. The background is dark and out of focus.

Können wir uns die Hände in
Unschuld waschen?

Eher nicht.

NO



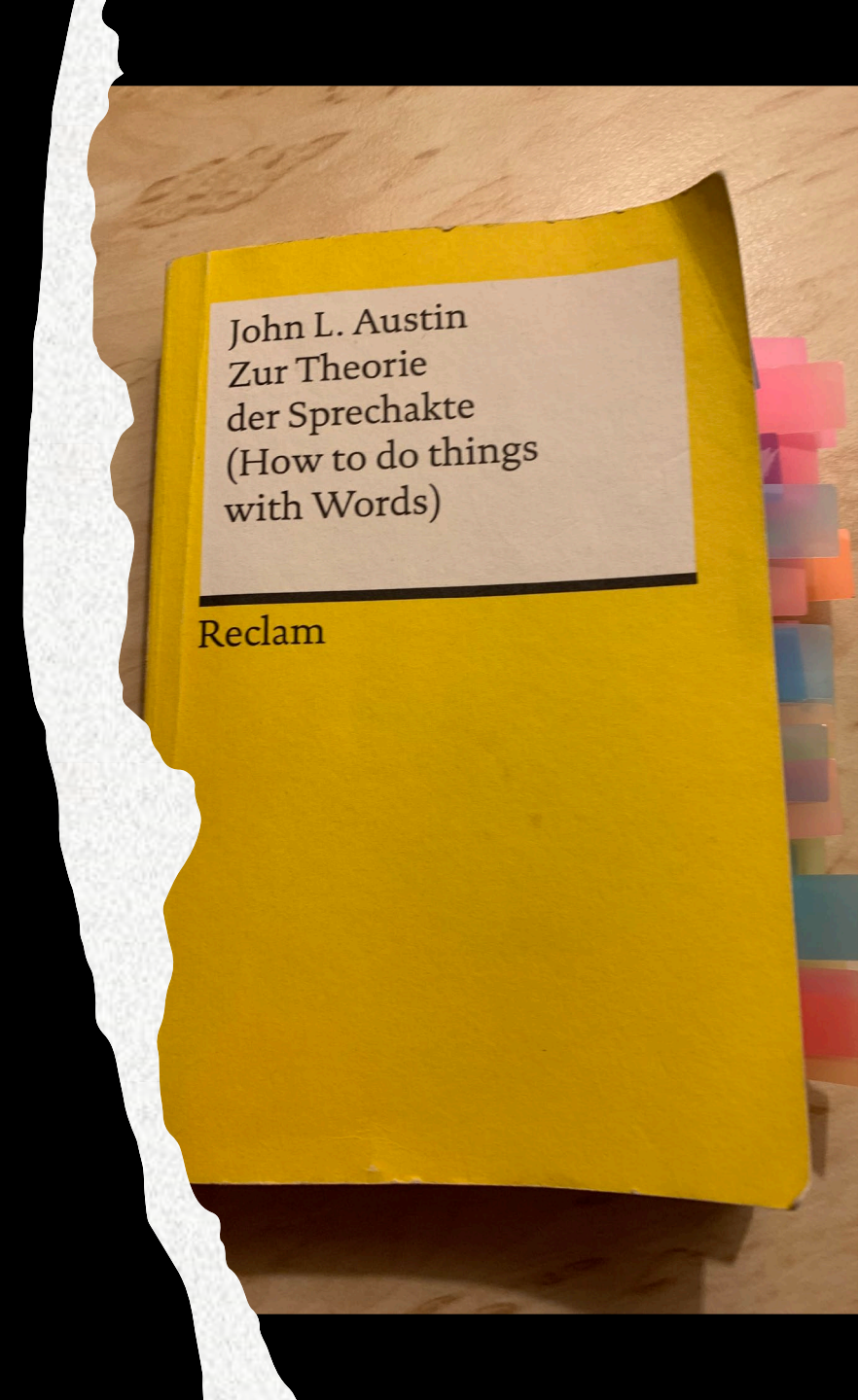
AI Act erwähnt „trustworthy AI“ über 15 Mal.



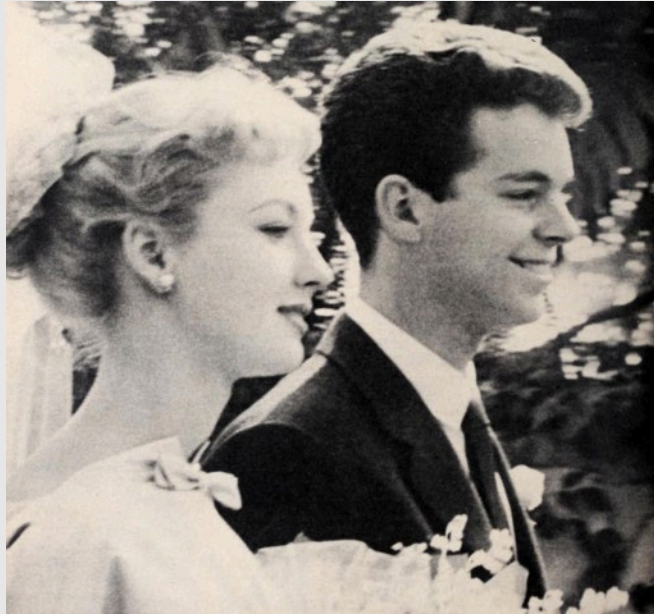
Robert Williams, festgenommen in Detroit



Anwendung der
Sprechakttheorie
von Austin (1962)



Beispiele für Sprechakte



Äußerungen, die
“Fakten” schaffen, an die
sich die betreffenden
Personen (und u.U. die
Gesellschaft als Ganzes)
danach halten sollen.

Sprechakttheorie

Damit Sprechakte gelingen, müssen „Gelingensbedingungen“ (*felicity conditions*) eingehalten werden:

A1: (...) Befugnis der Beteiligten.

A2: Personen, Gegenstände und Umstände müssen passen.

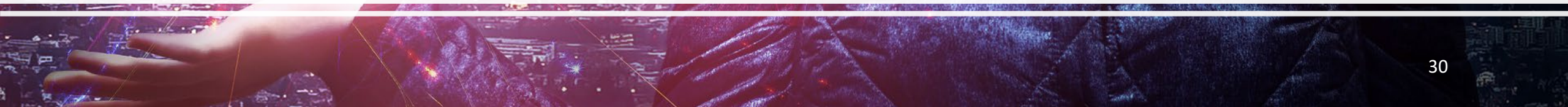
B1 & B2: Alle Beteiligten müssen den Vorgang korrekt und vollständig durchführen.

Γ 1: Der Sprechakt darf nicht unglaubwürdig oder unaufrichtig sein.

Γ 2: (...)



Bewertungen und Vorhersagen sind Sprechakte



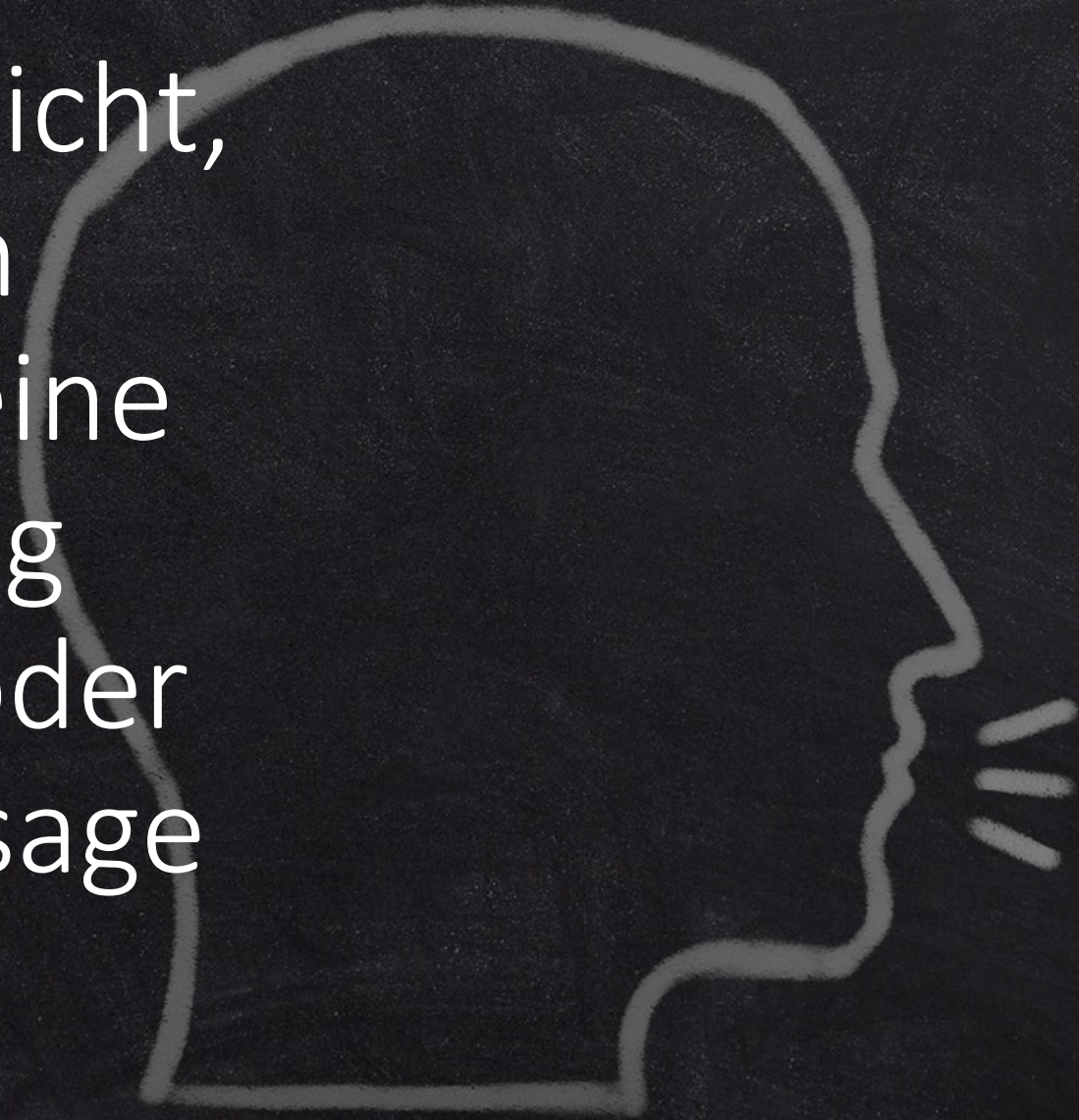
Heuristiken: Wenn es keine schnellen Algorithmen gibt

Eine Heuristik ist eine Abfolge von Aktionen, die zu einer Lösung eines Problems führen kann, aber keine Garantie dafür bietet, wie weit diese Lösung von der optimalen Lösung entfernt ist.



Bsp: Rechte-Hand-Regel im Labyrinth.

Aber wer spricht,
wenn ein
Computer eine
Bewertung
berechnet oder
eine Vorhersage
trifft?



Wir sprechen.
Translatiert in Zeit und Raum.





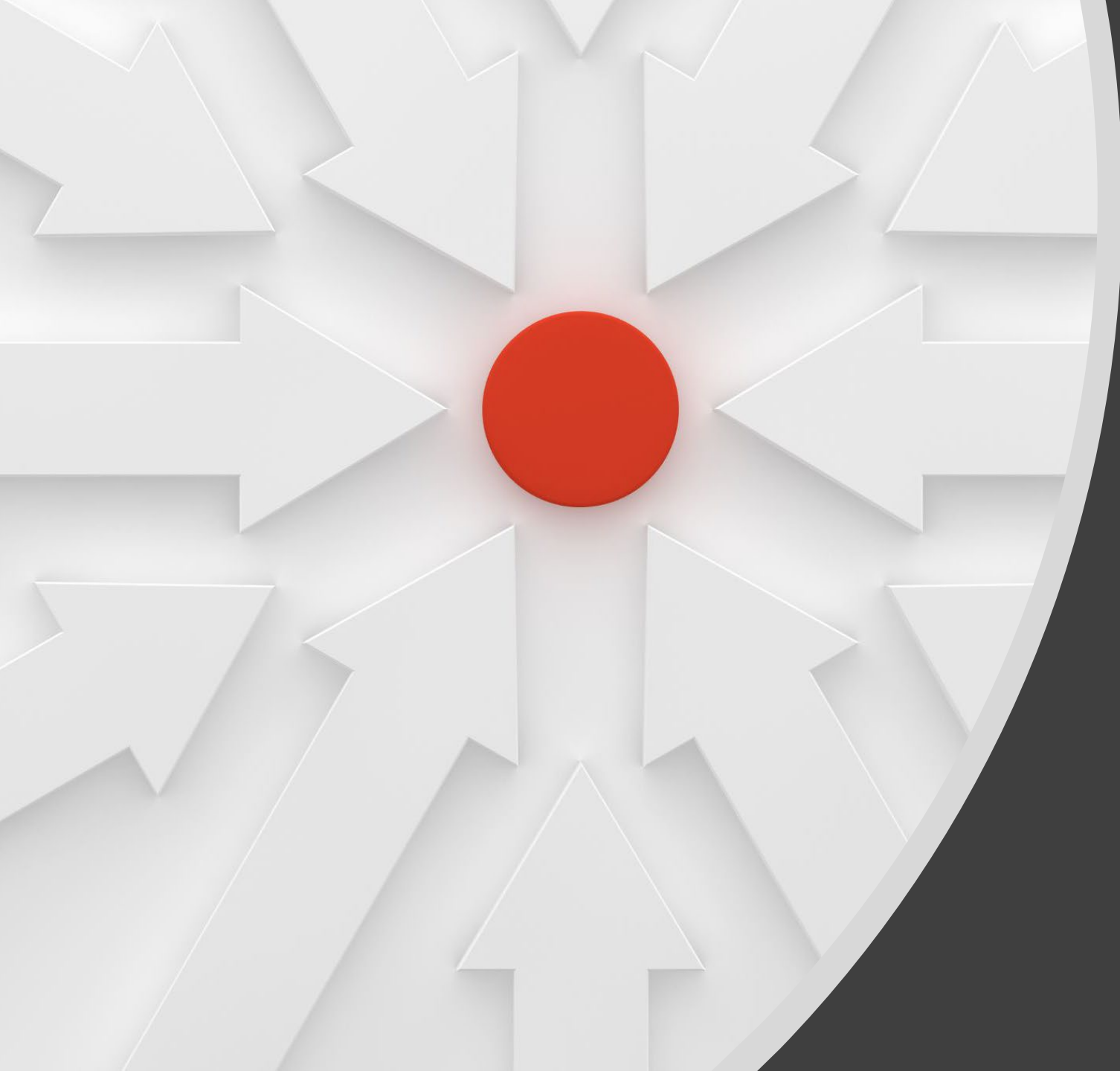
Der
Sprechakt
und
Technologie



Robert Williams, festgenommen in Detroit

```
mirror_mod = modifier_ob.  
set mirror object to mirror.  
mirror_mod.mirror_object  
operation == "MIRROR_X":  
mirror_mod.use_x = True  
mirror_mod.use_y = False  
mirror_mod.use_z = False  
operation == "MIRROR_Y":  
mirror_mod.use_x = False  
mirror_mod.use_y = True  
mirror_mod.use_z = False  
operation == "MIRROR_Z":  
mirror_mod.use_x = False  
mirror_mod.use_y = False  
mirror_mod.use_z = True  
  
#selection at the end -add  
mirror_ob.select= 1  
modifier_ob.select=1  
context.scene.objects.active  
("Selected" + str(modifier_ob.  
mirror_ob.select = 0  
= bpy.context.selected_object  
data.objects[one.name].select  
  
print("please select exactly  
  
-- OPERATOR CLASSES ----  
  
types.Operator):  
on X mirror to the selected  
object.mirror_mirror_x"  
mirror X"  
  
context):  
context.active_object is not
```

Hypothese 1: Wir sind für die genaue Formulierung des Outputs als Sprechende verantwortlich, damit der Sprechakt gelingen kann.



Hypothese 2:
Verwender sind für
die Verwendbarkeit
der Methode für
ihre Zwecke
verantwortlich.



Hypothese 3: Dazu müssen wir bestmöglich transparent machen, wann eine Methode aus technischer Sicht und basierend auf ihren Modellannahmen verwendbar ist.

Zusammenfassung

A wide-angle photograph of a long, straight asphalt road stretching towards the horizon. The road has a white dashed center line and is flanked by fields of dry, golden-brown grass. In the distance, there are some trees and low hills under a dramatic sky with large, dark, grey clouds and patches of blue. The overall mood is contemplative and open.

Data Science + Ethik

- Unsere Algorithmen und Heuristiken sind sehr komplex geworden.
- Insbesondere beim maschinellen Lernen gibt es emergente Phänomene.
- Hypothese 1: Wir begehen einen Sprechakt, wenn unsere Methode auf Daten angewendet wird.
 - Wir sind für ihn verantwortlich.
 - Daher muss der Output sorgfältig formuliert werden.
- Hypothese 2: Verwender unserer Methode sind für die ordnungsgemäße Verwendung verantwortlich.
 - Hypothese: Verwender müssen sich in die ordnungsgemäße Nutzung einlesen.
- Hypothese 3: Wir müssen unsere Methoden sorgfältig dokumentieren und aufklären, wann sie inhaltlich sinnvoll verwendet werden können.

A pair of black leather shoes and a green patterned tie are placed on a woven basket. The shoes are in the background, and the tie is in the foreground. The basket is made of brown woven material.

Fragen?
