



Gemälde von Sarina Walter  
mit freundlicher Genehmigung der Künstlerin

# Beeinflussung durch **Künstliche Intelligenz**

Über die Banalität  
der Beeinflussung  
und das Leben  
mit Algorithmen



Kate Crawford

Die Menschen befürchten, dass Computer zu intelligent werden und die Welt erobern.

**Aber Computer sind zu dumm und haben die Welt bereits erobert.**

# Über uns



**KAREN ULLRICH**

karenullrich.info

@karen\_ullrich

**AMSTERDAM MACHINE  
LEARNING LAB  
UNIVERSITEIT VAN AMSTERDAM**



**HENDRIK HEUER**

hen-drik.de

@hen\_drik

**INSTITUT FÜR  
INFORMATIONSMANAGEMENT BREMEN  
UNIVERSITÄT BREMEN**



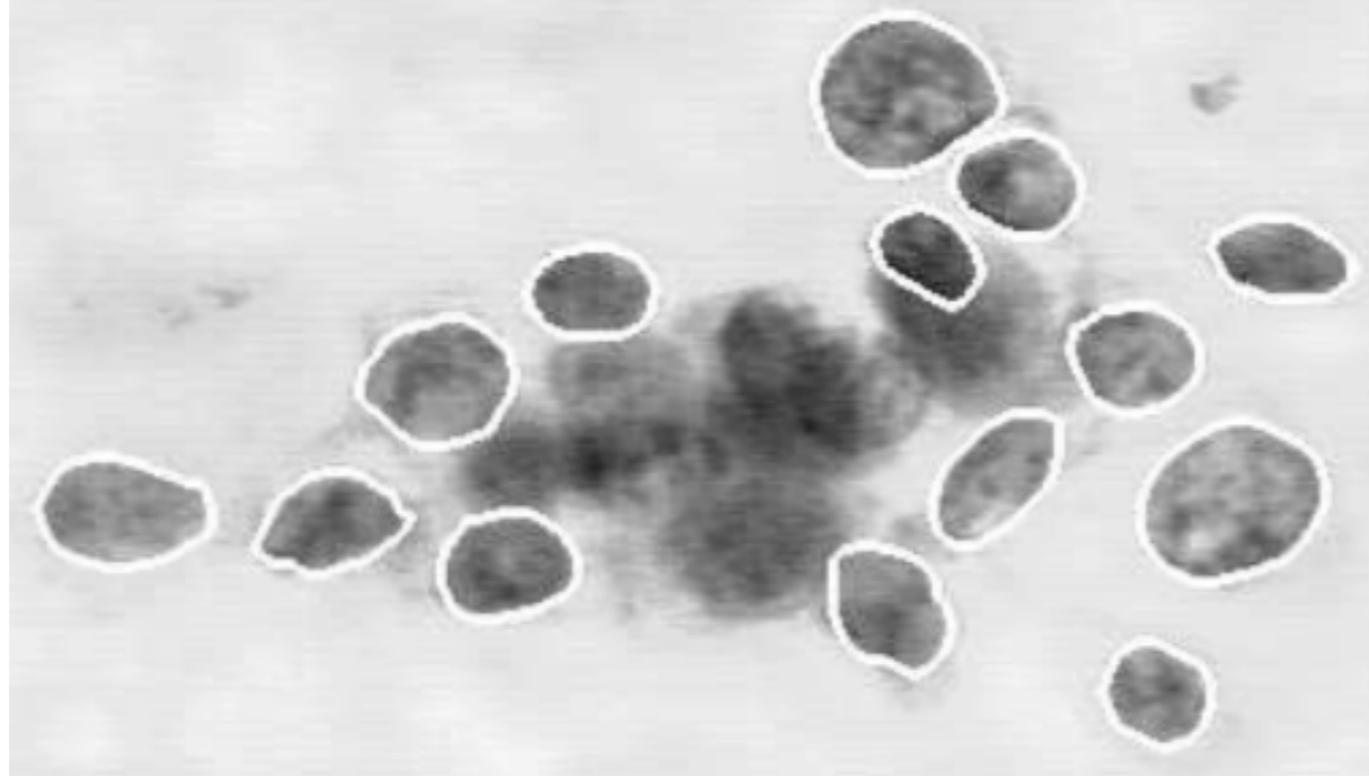
# Spam erkennen

## IMPERATIVE PROGRAMMIERUNG

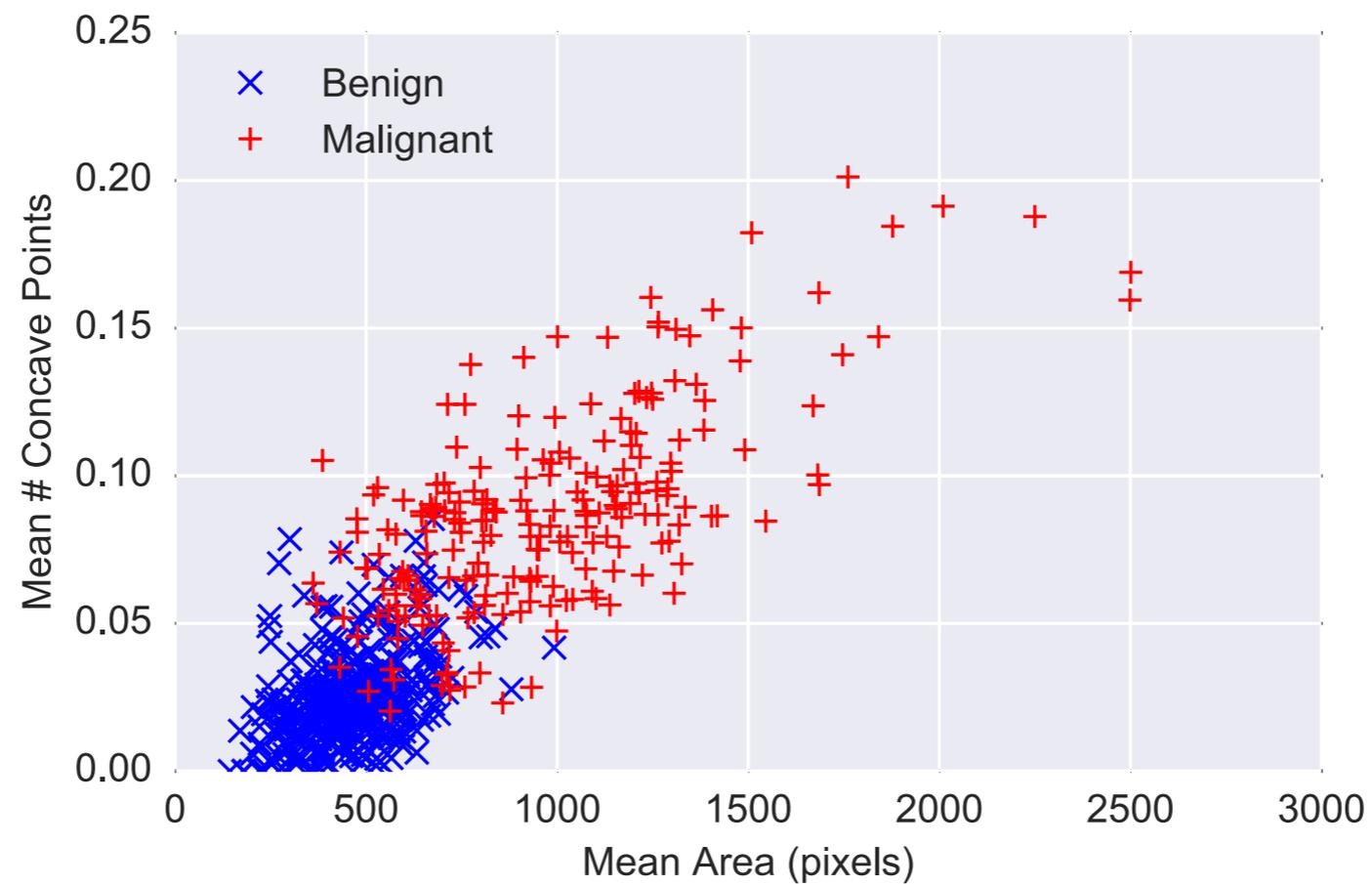
```
if "V!agrå" in mail:  
    return "Spam"  
else:  
    return "Ham"
```

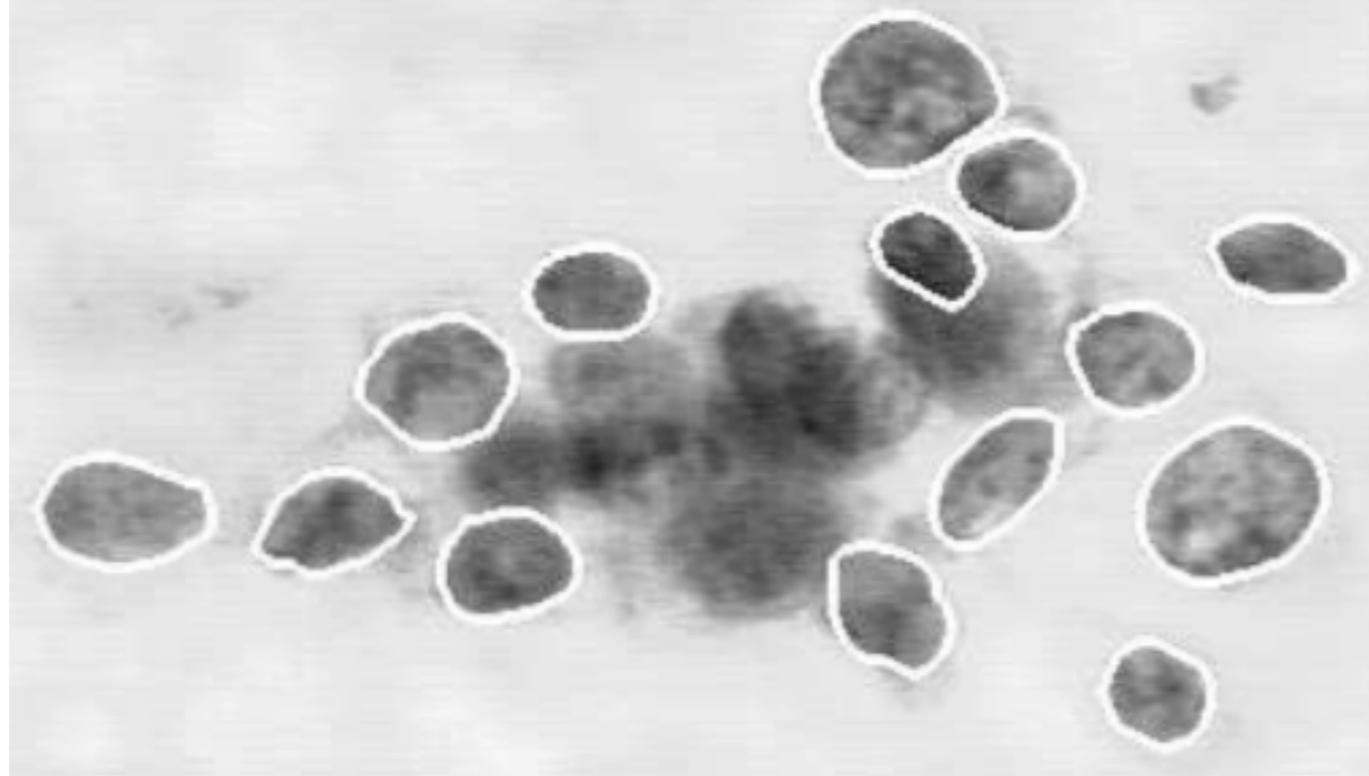
## MASCHINELLES LERNEN

1. prognostiziere für viele Beispiele, ob Nachricht Spam ist
2. berechne Fehler
3. ändere Parameter, um Fehler zu verringern
4. goto 1

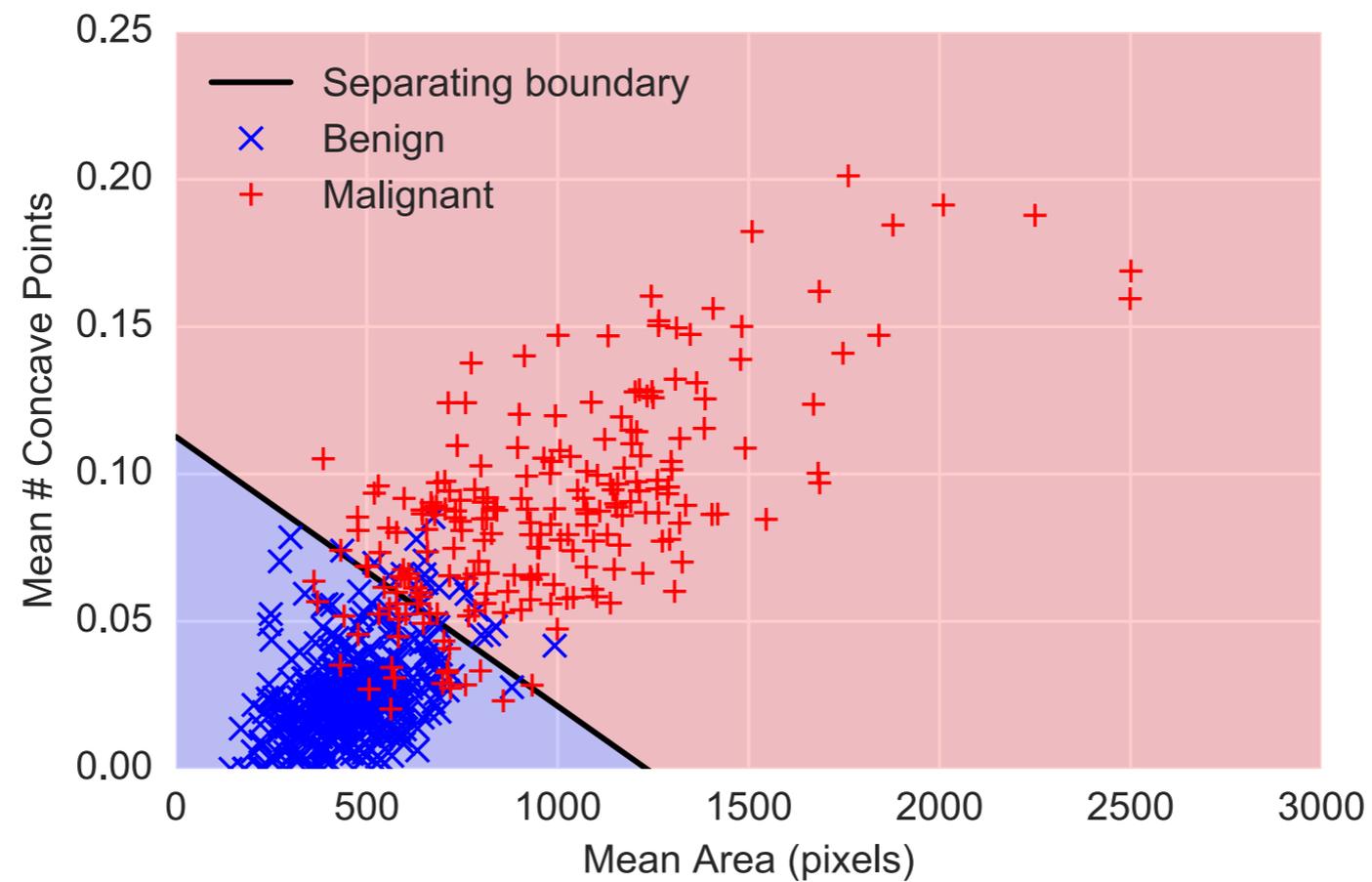


# Brustkrebs erkennen





# Brustkrebs erkennen



Kosinski et al.: Private traits  
and attributes are predictable  
from digital records of human  
behavior



# Persönlichkeitsmerkmale erkennen

Weißer/r oder Afroamerikaner/in: 95%

Männlich oder Weiblich: 93%

Heterosexuell oder homosexuell: 88%

Demokrat oder Republikaner: 85%

Christlich oder muslimisch: 82%

Drogenerfahrung: 65% - 73%

Ehe der Eltern hat bis zum 21. Lebensjahr gehalten: 60%

Kosinski et al.: Private traits and attributes are predictable from digital records of human behavior



# Persönlichkeitsmerkmale erkennen







LOW RISK

HIGH RISK

$x^i$

$y^i$

Deutschlandfunk:  
Algorithmen im US-  
Justizsystem  
Schicksalsmaschinen  
von Tom Schimmeck

# Verbrechen vorhersagen



Risk Scales

MOOC Non-VIO Risk	High
MOOC VIO Risk	High
Scale	1 2 3 4 5 6 7 8 9

Criminogenic Need Scales

Criminal Involvement	General Involvement	High
	Status of Non-Compliance	High
	History of Violence	High
	Current Violence	Low
Relationships/Community	General Association/Peers	Problematic
	Criminal Opportunity	Problematic
	Leisure and Recreation	Problematic
	Social Isolation	Problematic
	Substance Abuse	Problematic
Personality/Attitudes	Criminal Personality	Highly Problematic
	Cognitive Behavioral	Problematic
Family	Family Cohesivity	Problematic
Social Exclusion	Neighborhood/Education	Problematic
	Residential Instability	Problematic
	Social Stigmatization	Problematic







English Spanish French German - detected



English Spanish Turkish

Translate

Er ist ein Babysitter.  
Sie ist eine Ärztin.



43/5000

O bir bebek bakıcısıdır.  
O bir doktor.



English Spanish Turkish Detect language



English German Turkish

Translate

O bir bebek bakıcısıdır.  
O bir doktor.



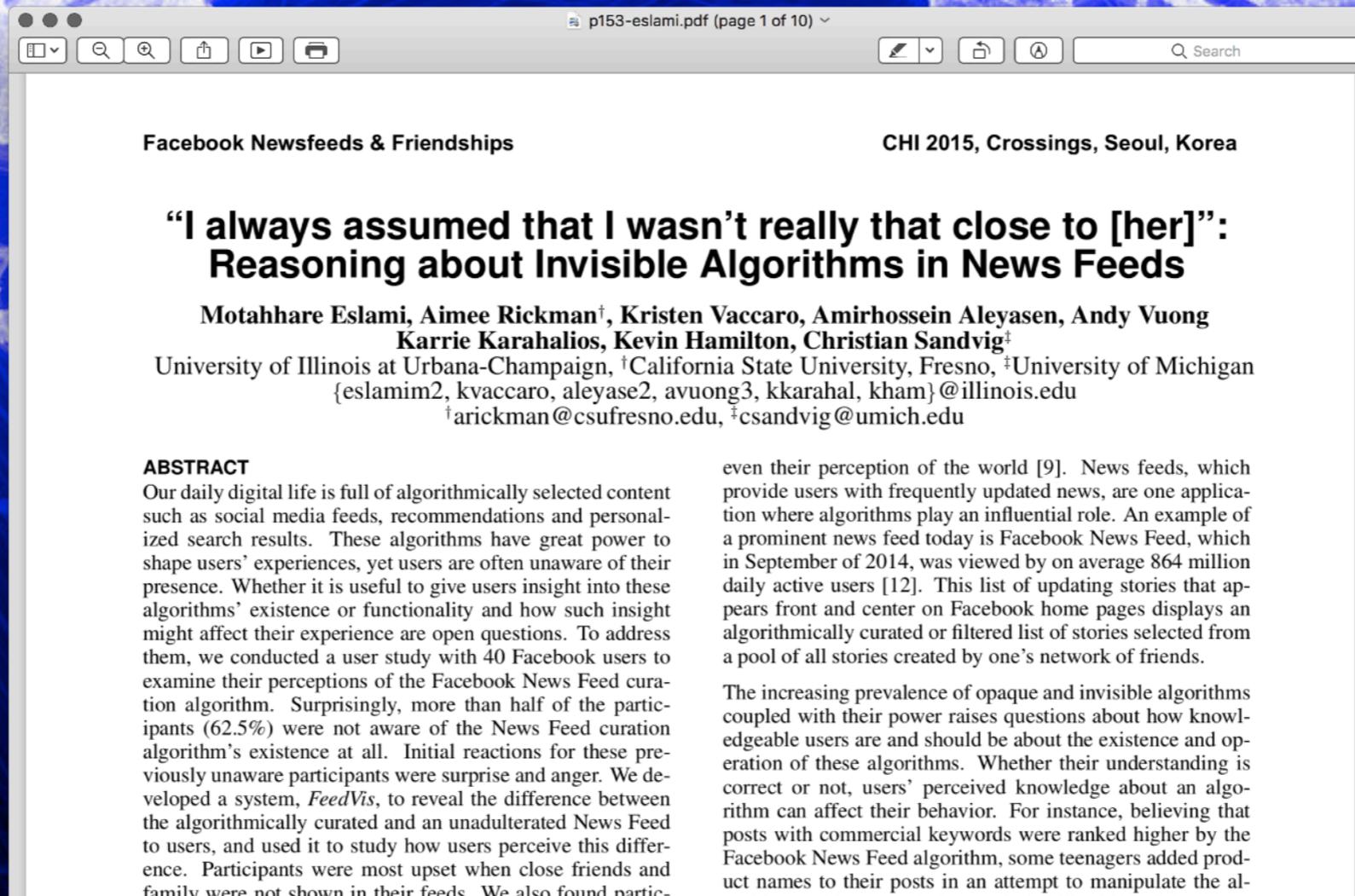
38/5000

Sie ist eine Babysitterin.  
Er ist Arzt.





# **Algorithmic** News Curation



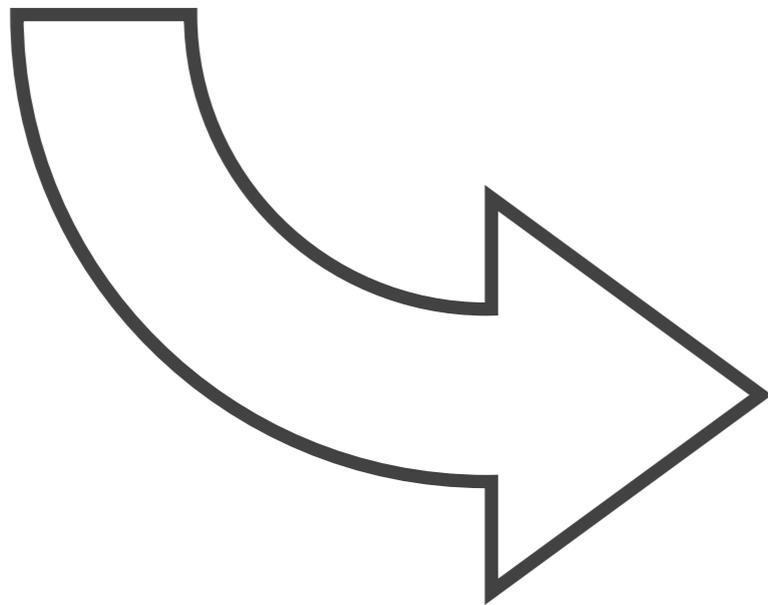
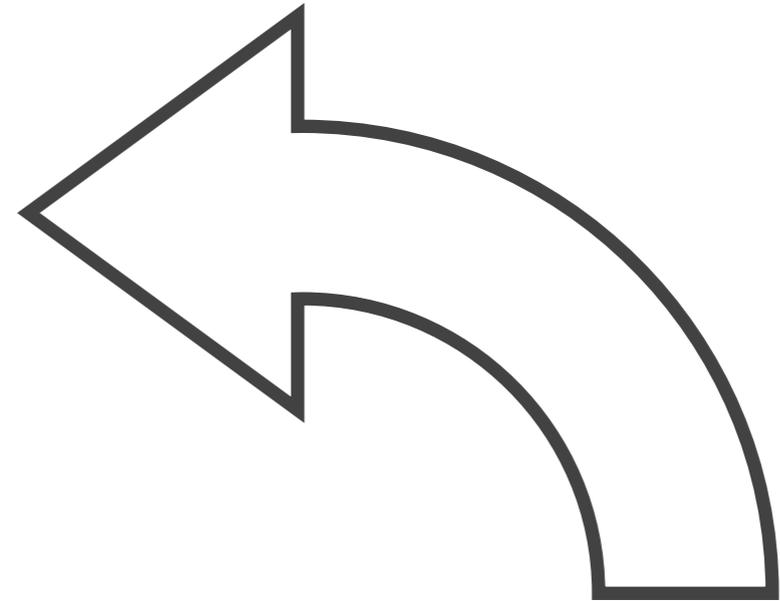
# Algorithmic News Curation



**Gemälde von Sarina Walter**  
mit freundlicher Genehmigung der Künstlerin

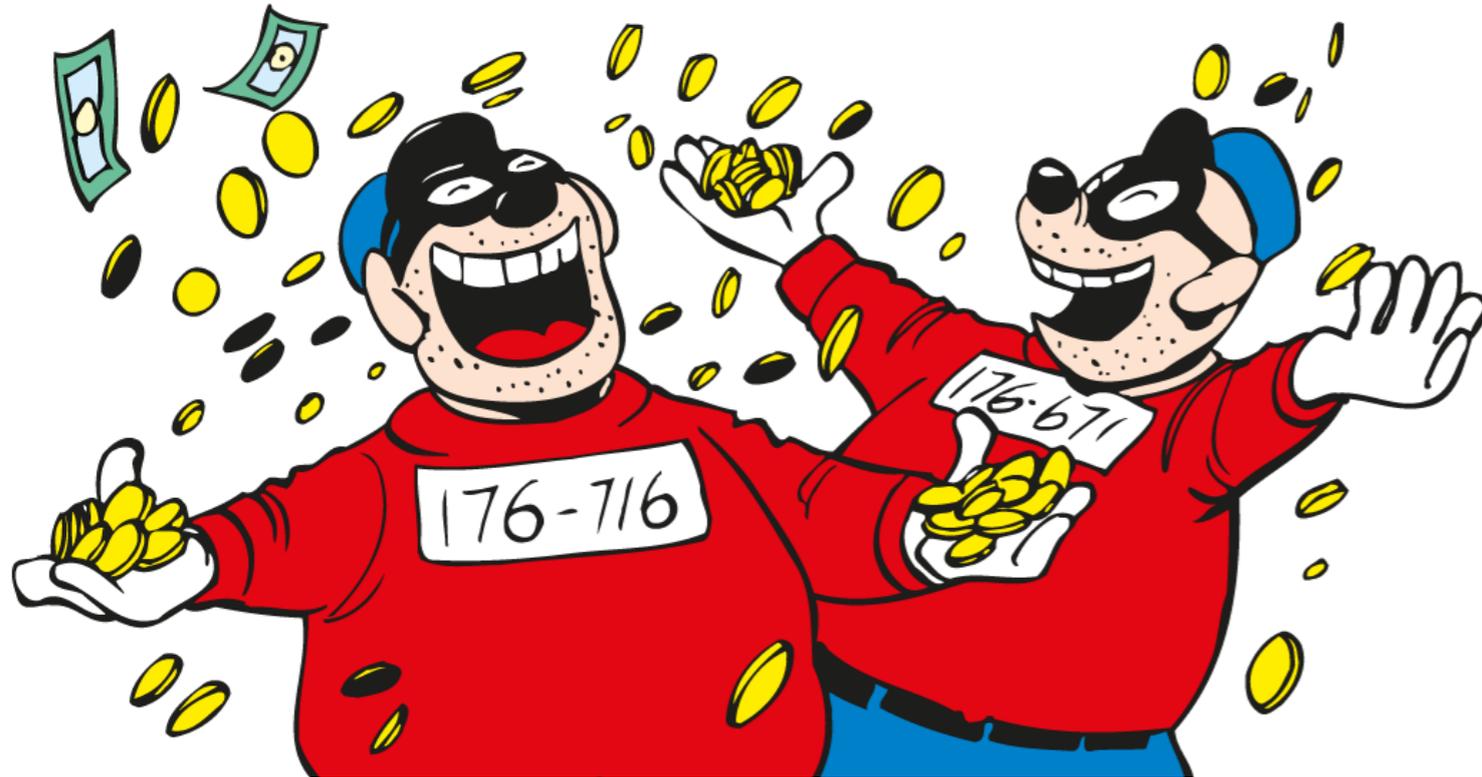
Wieviel Vertrauen  
ist angebracht?  
Wann scheitern  
die Algorithmen?

**Bias der Daten**



**Bias des Modells**

1800 Fotos, Chinesen, 18-55, keine Gesichtshaare, Narben, Tattoos



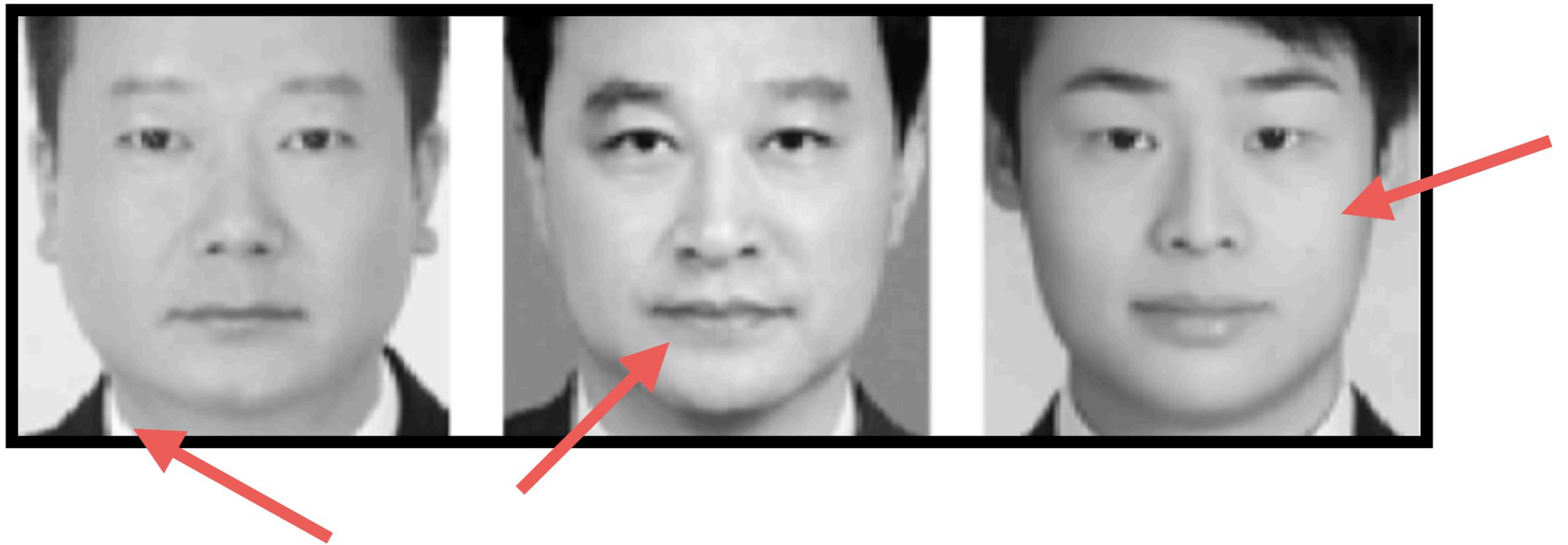
## Kriminelle

700 Fotos verurteilter Krimineller,  
bereitgestellt durch die Polizei.



## Rechtschaffende

1100 Fotos aus verschiedenen  
Quellen des Internets.



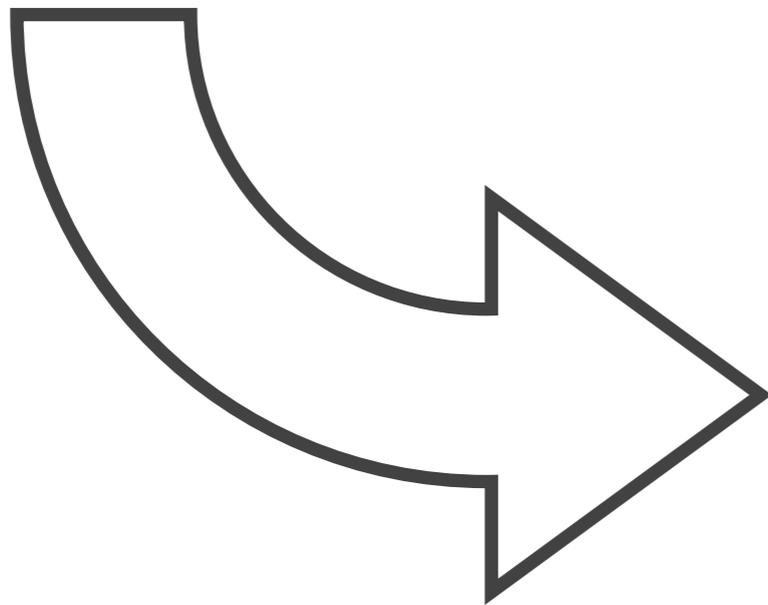
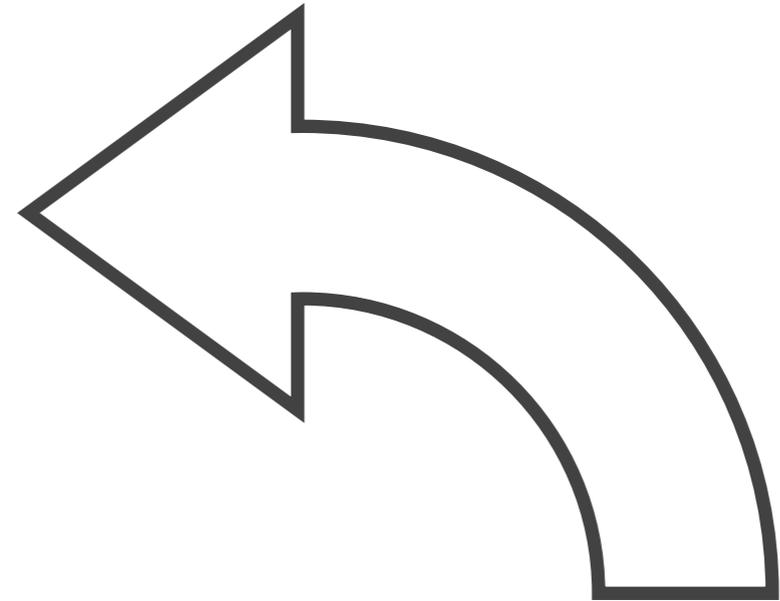
Verzerrung der Stichprobe

Anzahl der Daten

Melde-Bias

Ausgewählte Merkmale, die ein Problem beschreiben

**Bias der Daten**



**Bias des Modells**

## “No Free Lunch” - Theorem

Betrachtet man alle möglichen Situationen, funktioniert kein Modell besser als alle anderen.

Wir **müssen Annahmen machen**, um Szenarien zu beschreiben und Modelle zu vereinfachen (**Modell-Bias**).

**Aber:** Annahmen **versagen** in gewissen Situationen.

Unsere Aufgabe: Annahmen **benennen** und **überprüfen**, ob sie valide sind.

# **Zielsetzung:** Was ist Erfolg?

Ziele basieren auf Annahmen

Problem: Richtigkeit einer Übersetzung

Wort für Wort?

Paragraph für Paragraph?

Synonyme?

# **Zielsetzung:** Was ist Erfolg?

Ziele basieren auf Annahmen

Problem: Richtigkeit einer Übersetzung

**Satz für Satz**

**Sätze sind unabhängig**

**Nur das wahrscheinlichste Wort/**

**Phrase wird als richtig bewertet**

German ▾ Sie ist eine Frau. Sie ist Ärztin. <small>Edit</small>	Turkish ▾ O bir kadın. O bir doktor. <small>Edit</small>	German ▾ Sie ist eine Frau. Er ist Arzt.
--	---	---

# Zielsetzung als Bias

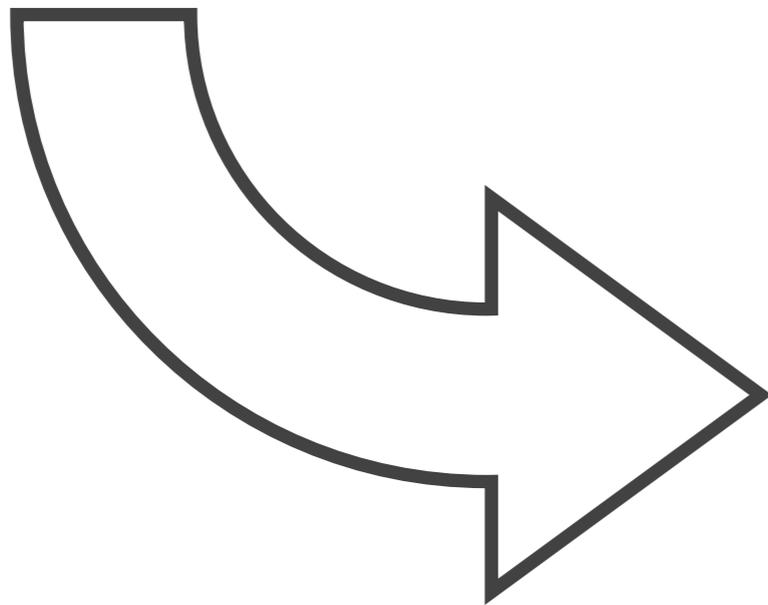
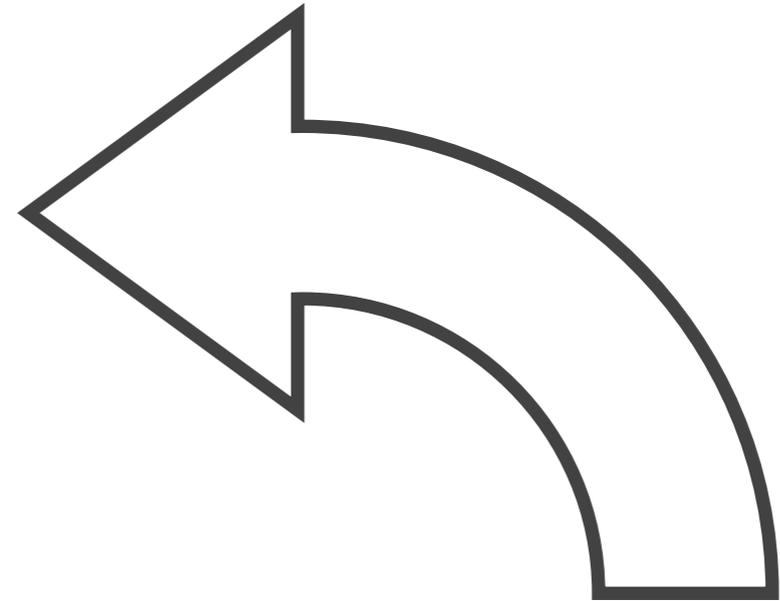
German ▾ Sie ist eine Frau, sie ist Ärztin. <small>Edit</small>	Turkish ▾ O bir kadın, o bir doktor. <small>Edit</small>	German ▾ Sie ist eine Frau, sie ist Ärztin.
--	---	--

[Open in Google Translate](#)

[Open in Google Translate](#)

[Feedback](#)

**Bias der Daten**

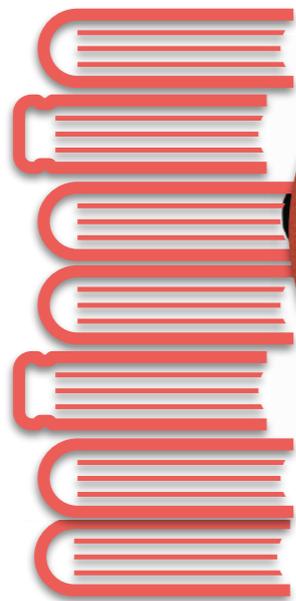
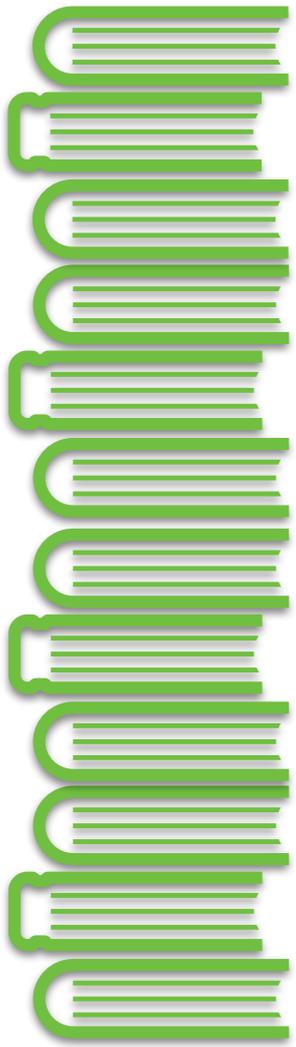


**Bias des Modells**

# Aktives Lernen

Human in the Loop

Anreicherung des  
Datensets



Bei unsicheren Daten



Algorithmus lernt durch  
Maschinelles Lernen





diri noir avec banan

@jackyalcine



Following

Google Photos, y'all [REDACTED] up. My friend's not a gorilla.



Skyscrapers



Airplanes



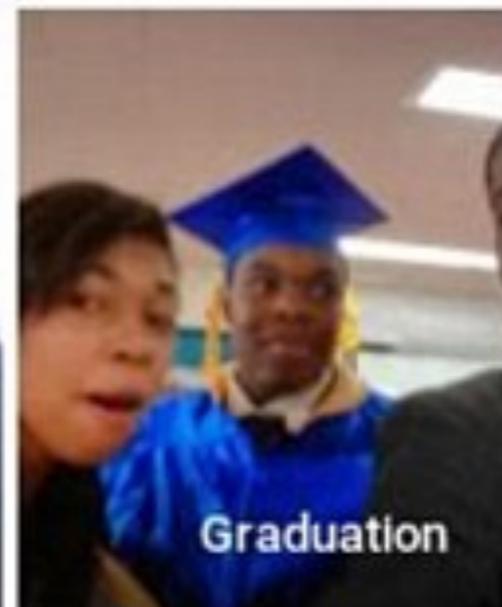
Cars



Bikes



Gorillas

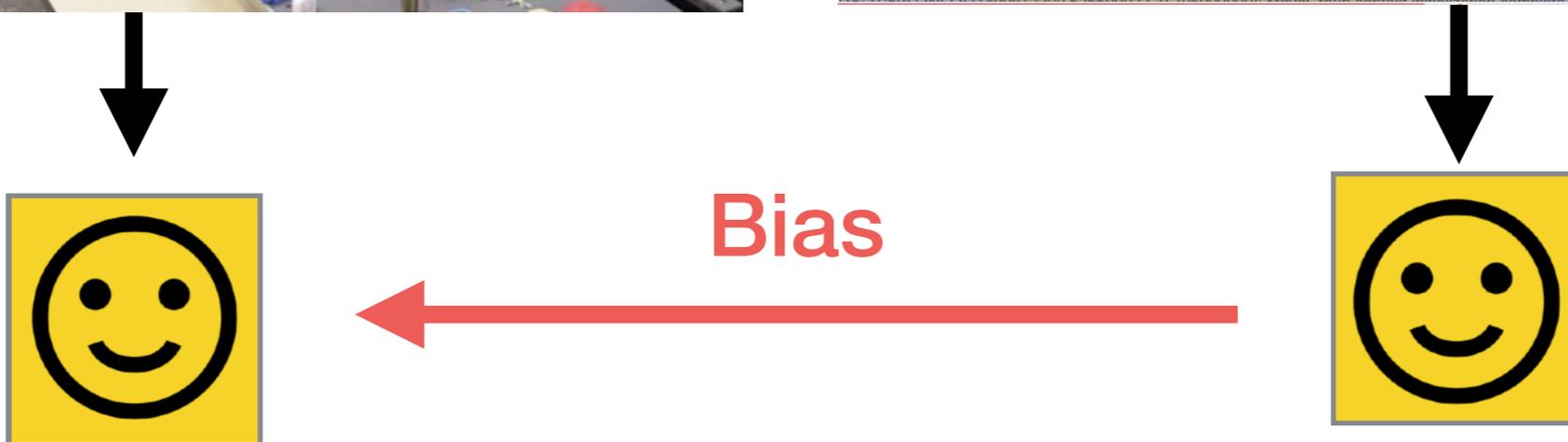
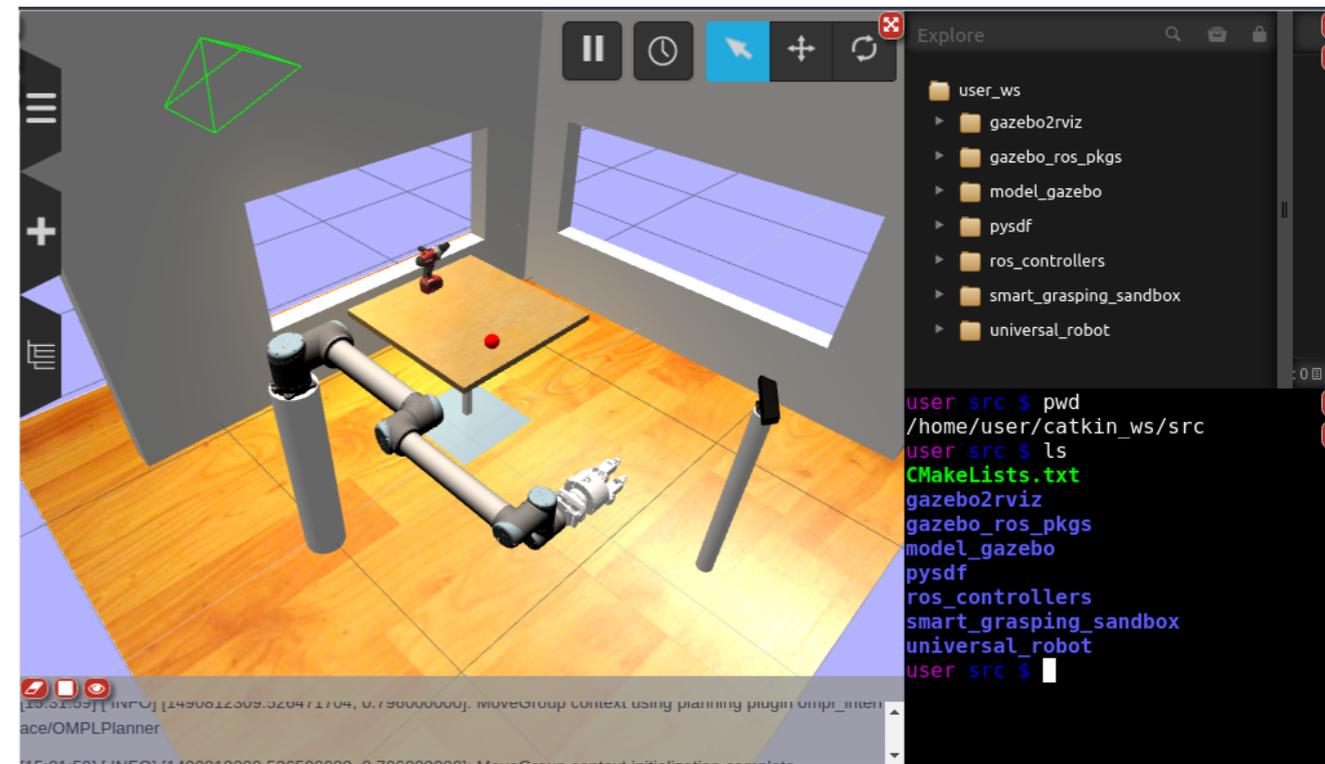


Graduation

# Übertragendes Lernen

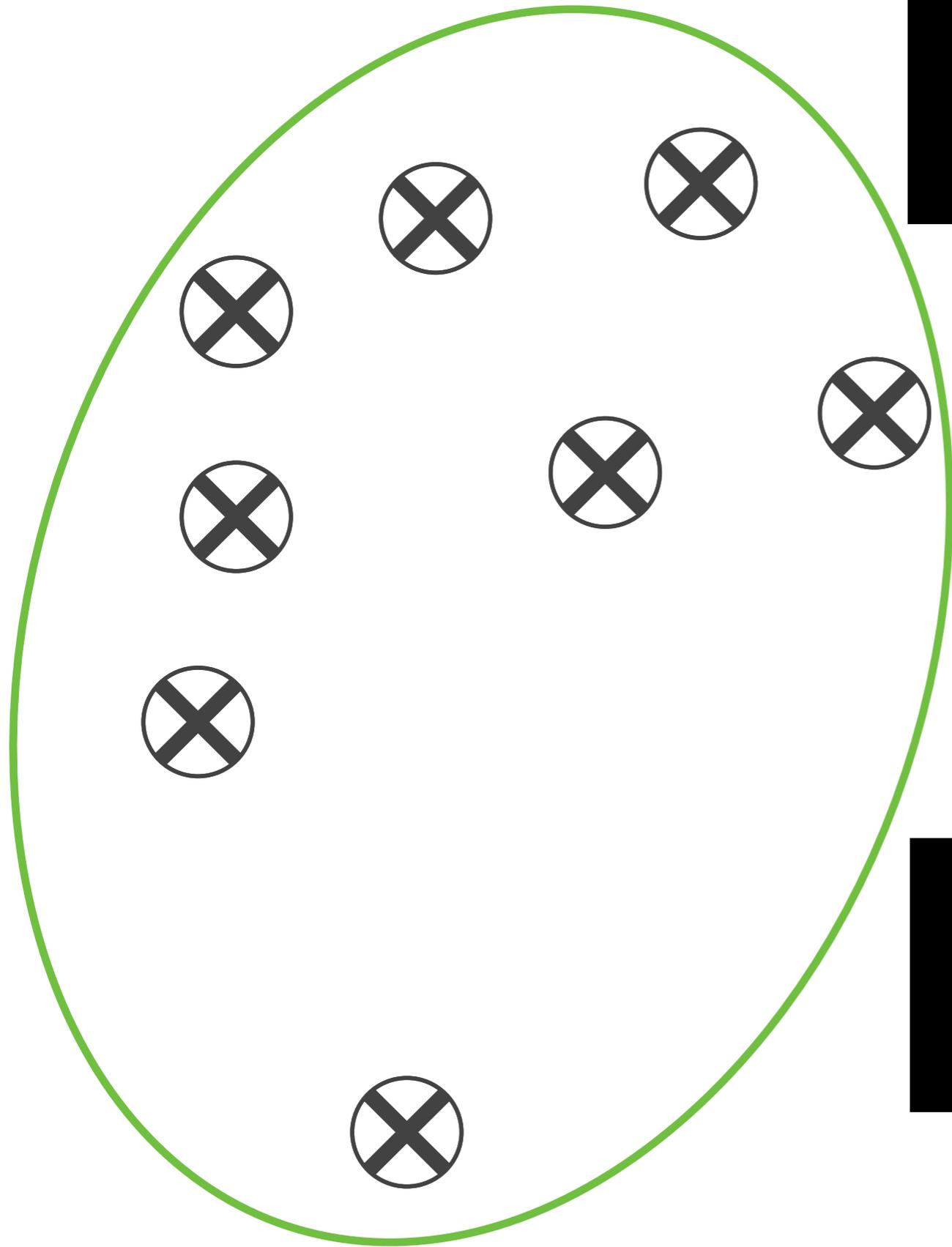
Ziel: Aufgabe erledigen  
Problem: Wenige Daten

Ähnliche Aufgabe, die  
erfolgreich gelöst werden kann





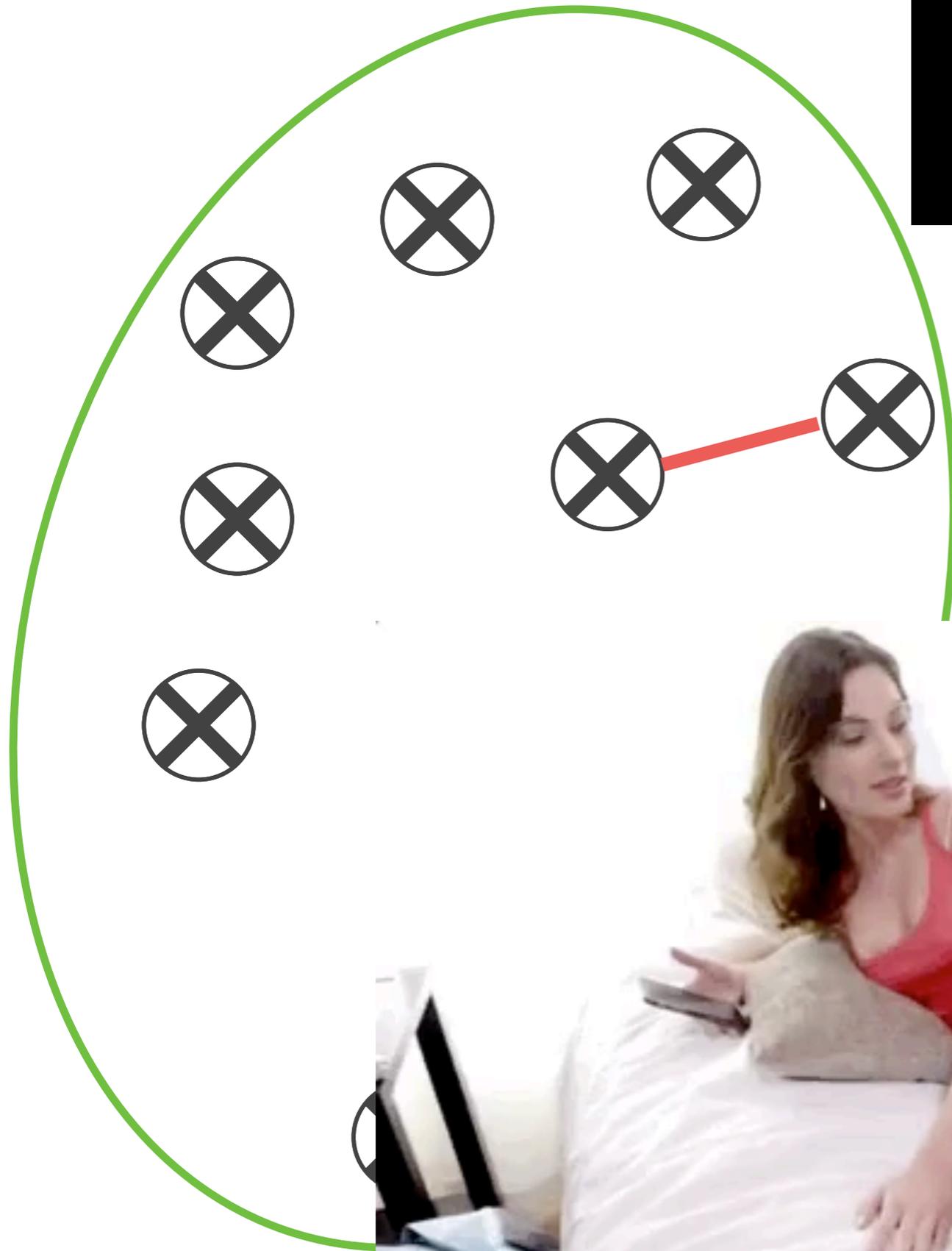
**Interpolieren**



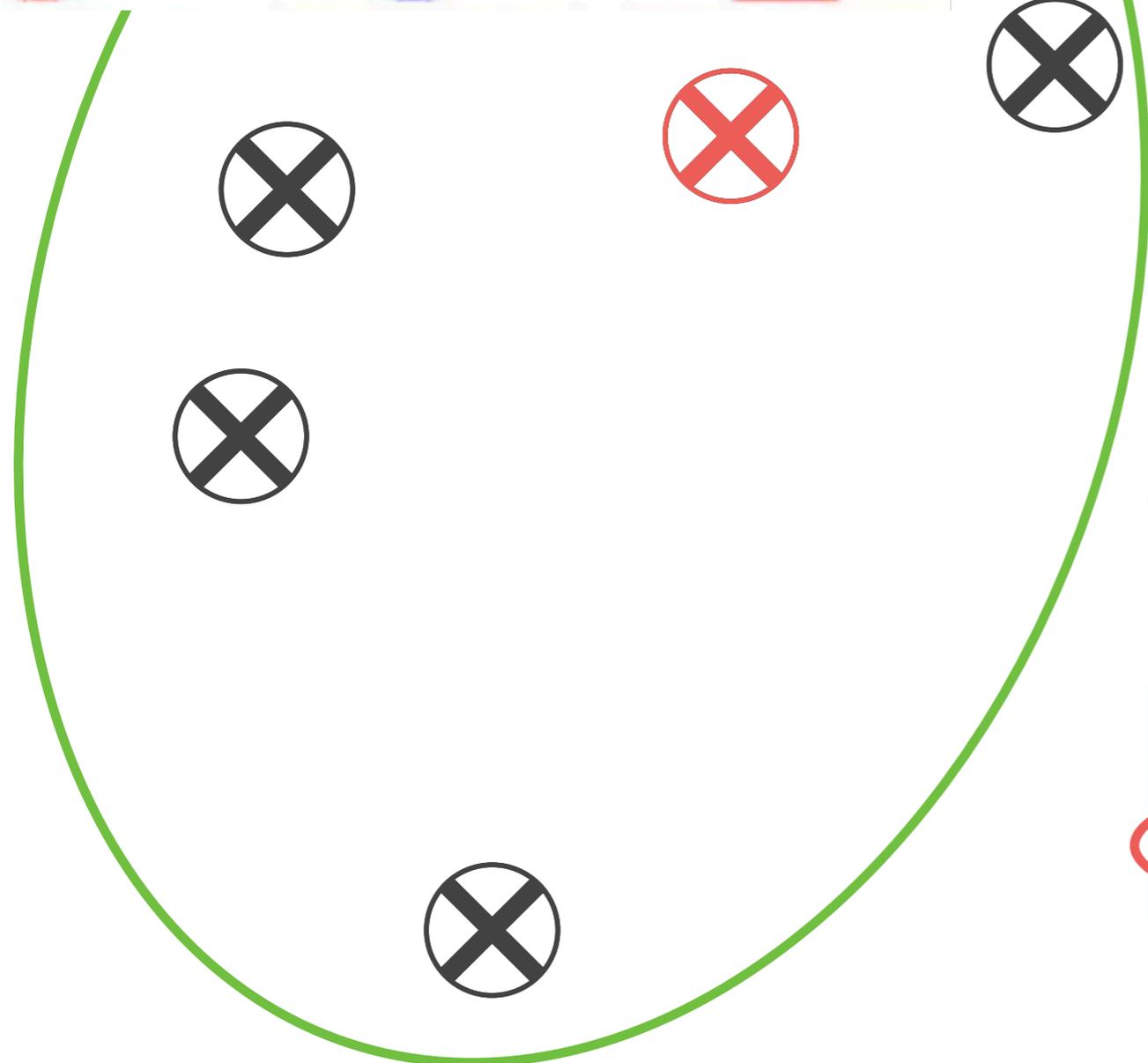
**Extrapolieren**

?

# Interpolieren



# Extrapolieren



?





Gemälde von Sarina Walter  
mit freundlicher Genehmigung der Künstlerin

# Beeinflussung durch **Künstliche Intelligenz**



**KAREN ULLRICH**

karenullrich.info  
@karen\_ullrich



**HENDRIK HEUER**

hen-drik.de  
@hen\_drik