„Maschinelles Lernen“

(Und warum das vielleicht gruseliger ist, als es klingt)

Jonas Betzendahl

@jbetzend

Wer bin ich?

Jonas Betzendahl
Masterstudiengang „Intelligente Systeme“
Technische Fakultät, Universität Bielefeld
jonas.betzendahl@gmail.com

Small Talk in Intelligent Systems

Die häufigste Frage an meinen Studiengang:

Small Talk in Intelligent Systems

Die häufigste Frage an meinen Studiengang:

- „Na, wie lange dauert es noch bis zur Roboterapokalypse?“


amazon

amazon

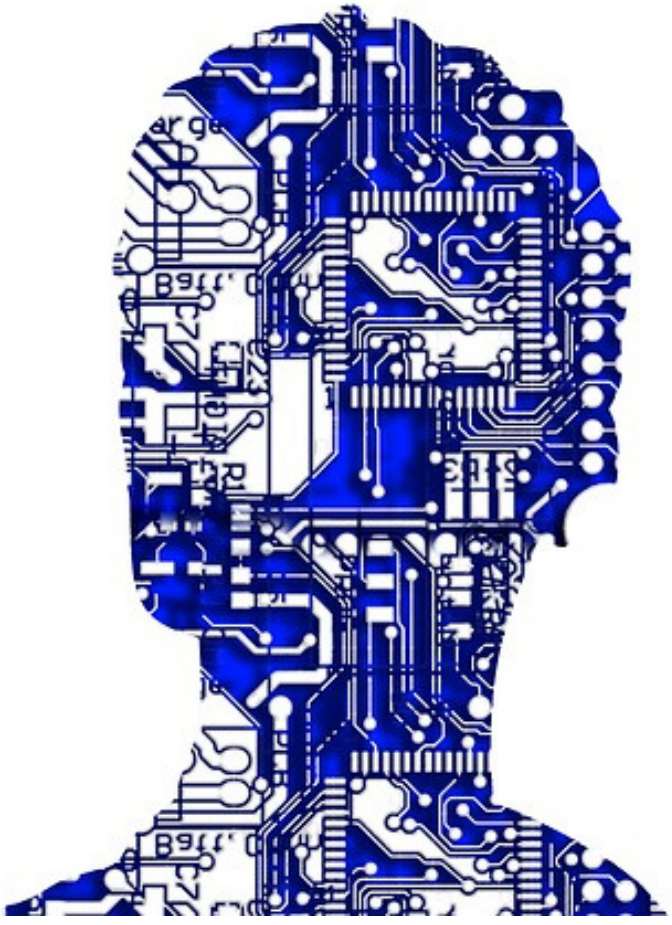

amazon
VANS

. . . zumindest habe ich bisher so immer meine Slams angefangen.

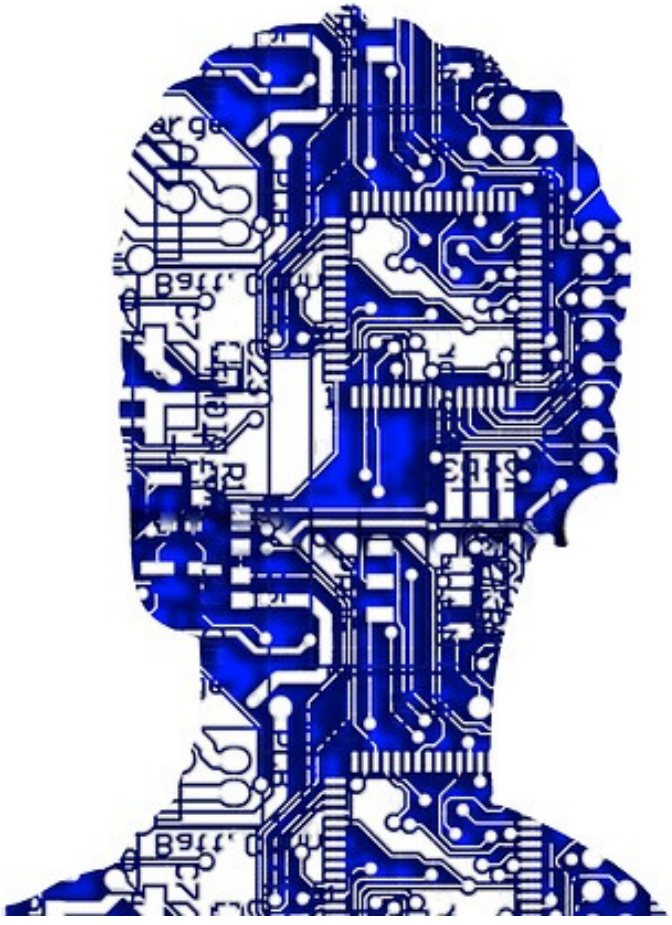
. . . zumindest habe ich bisher so immer meine Slams angefangen.

Wir müssen reden!

Wie funktioniert Maschinelles Lernen?

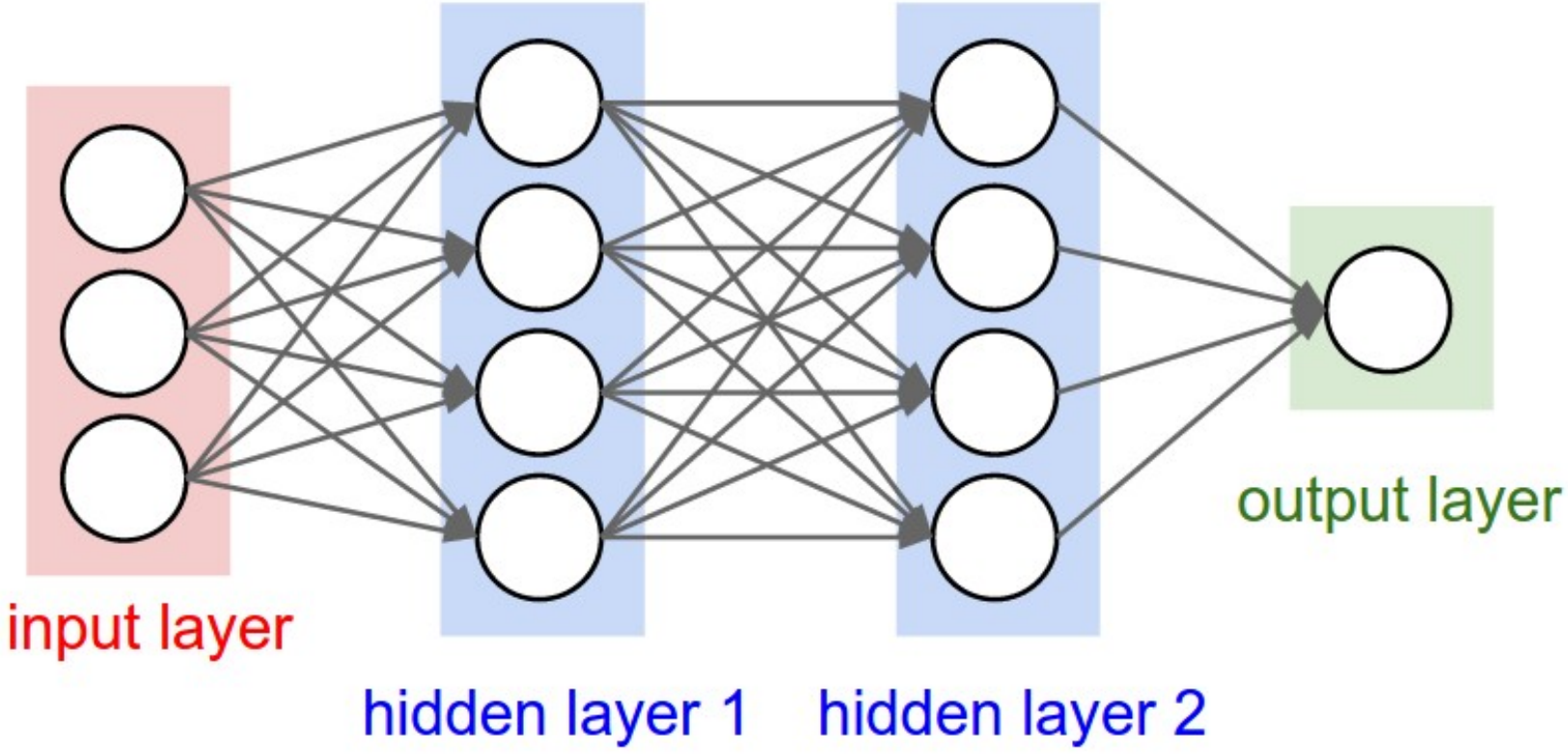


Maschinelles Lernen simuliert einen Vorgang nicht unähnlich dem im menschlichen Gehirn selbst.



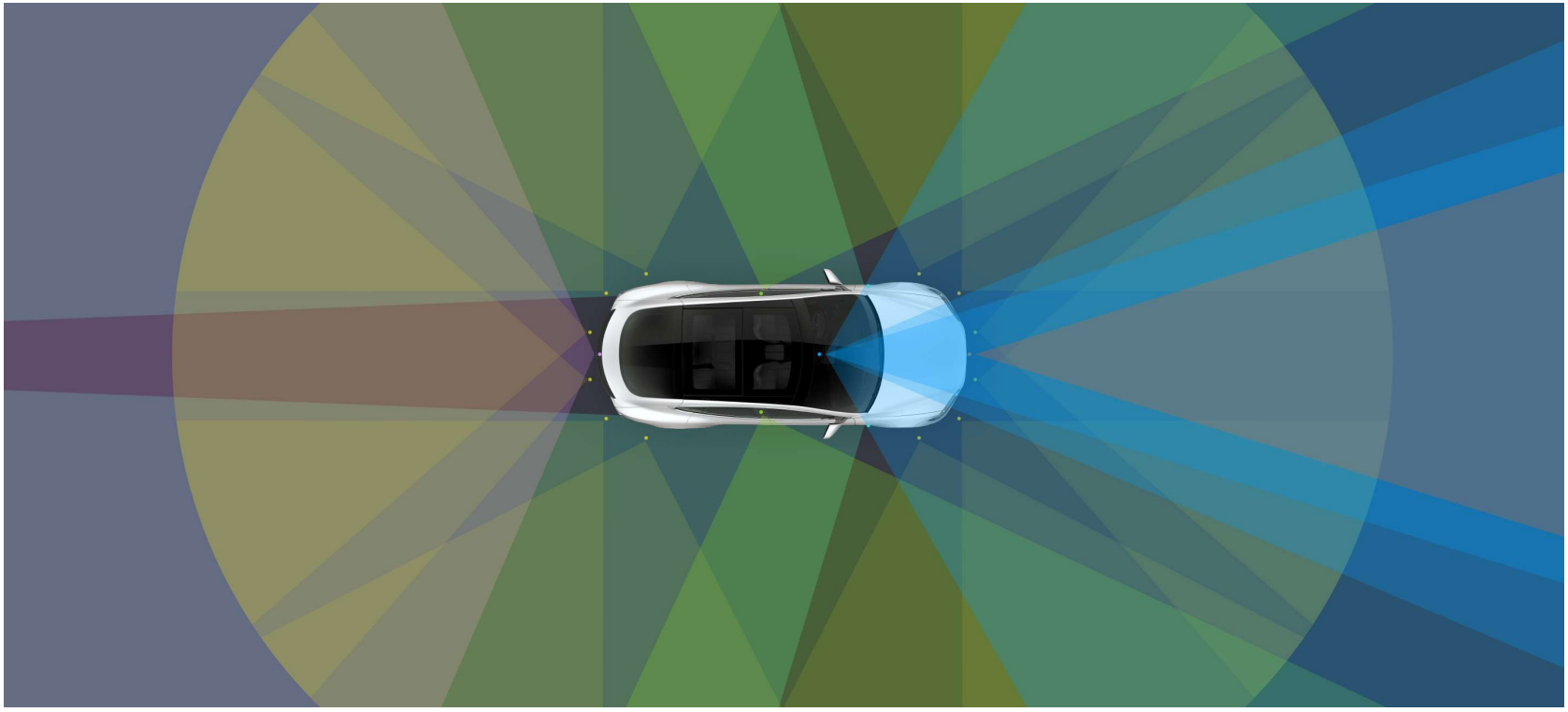
Maschinelles Lernen simuliert einen Vorgang nicht unähnlich dem im menschlichen Gehirn selbst.

Ein (künstliches) *neuronales Netz* wird simuliert und trainiert mit *Testdaten*, bis es akzeptable Leistungen bringt.


input layer
hidden layer 1
hidden layer 2
output layer

Die Errungenschaften von Maschinellem Lernen

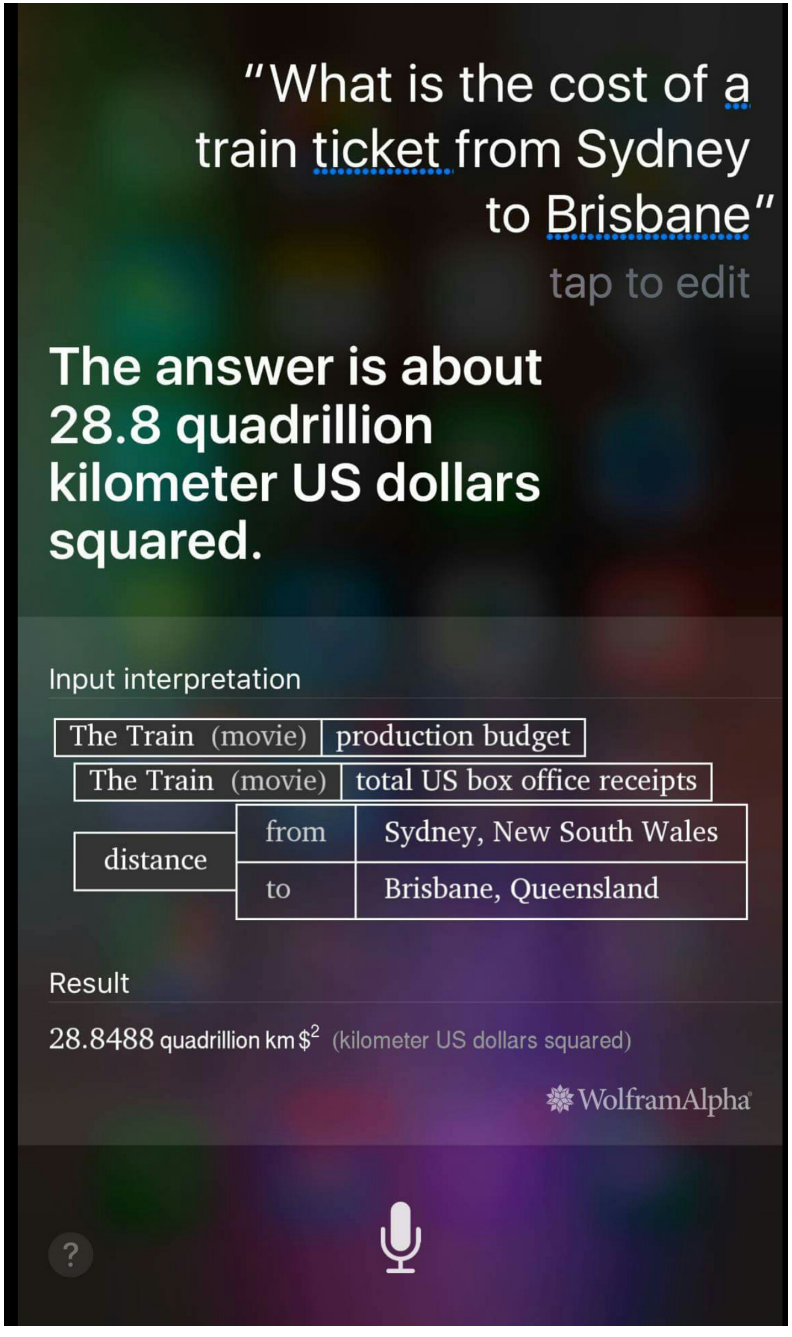
Maschinelles Lernen ist prinzipiell sehr mächtig und nützlich. . .



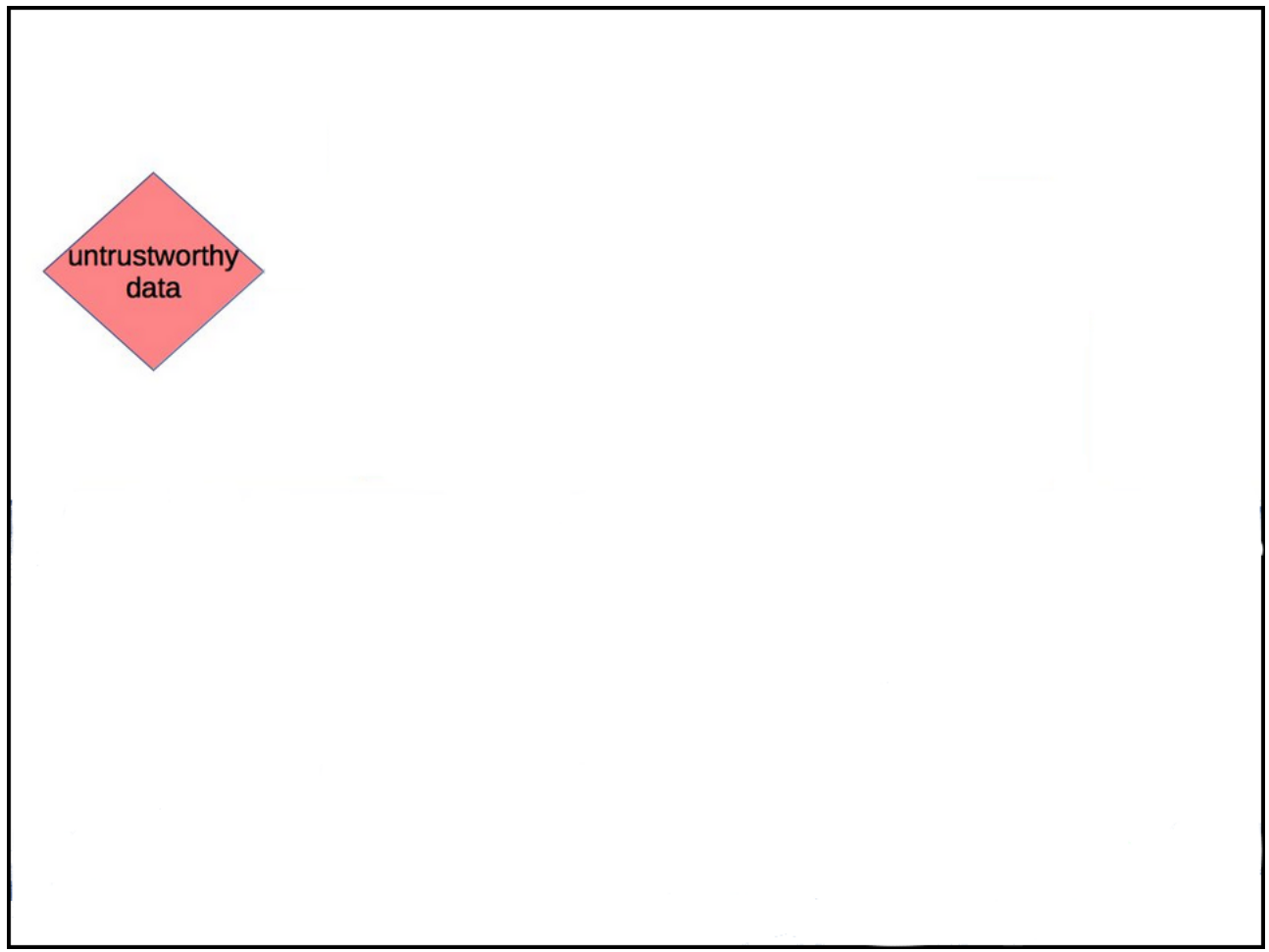
"Nachdem die Menschheit Jahrtausende damit verbracht hat, ihre Taktiken zu verbessern, erzählen uns die Computer, dass wir komplett daneben liegen.
Ich würde soweit gehen, zu sagen, dass noch kein einziger Mensch auch nur den Rand der Wahrheit von Go berührt hat."

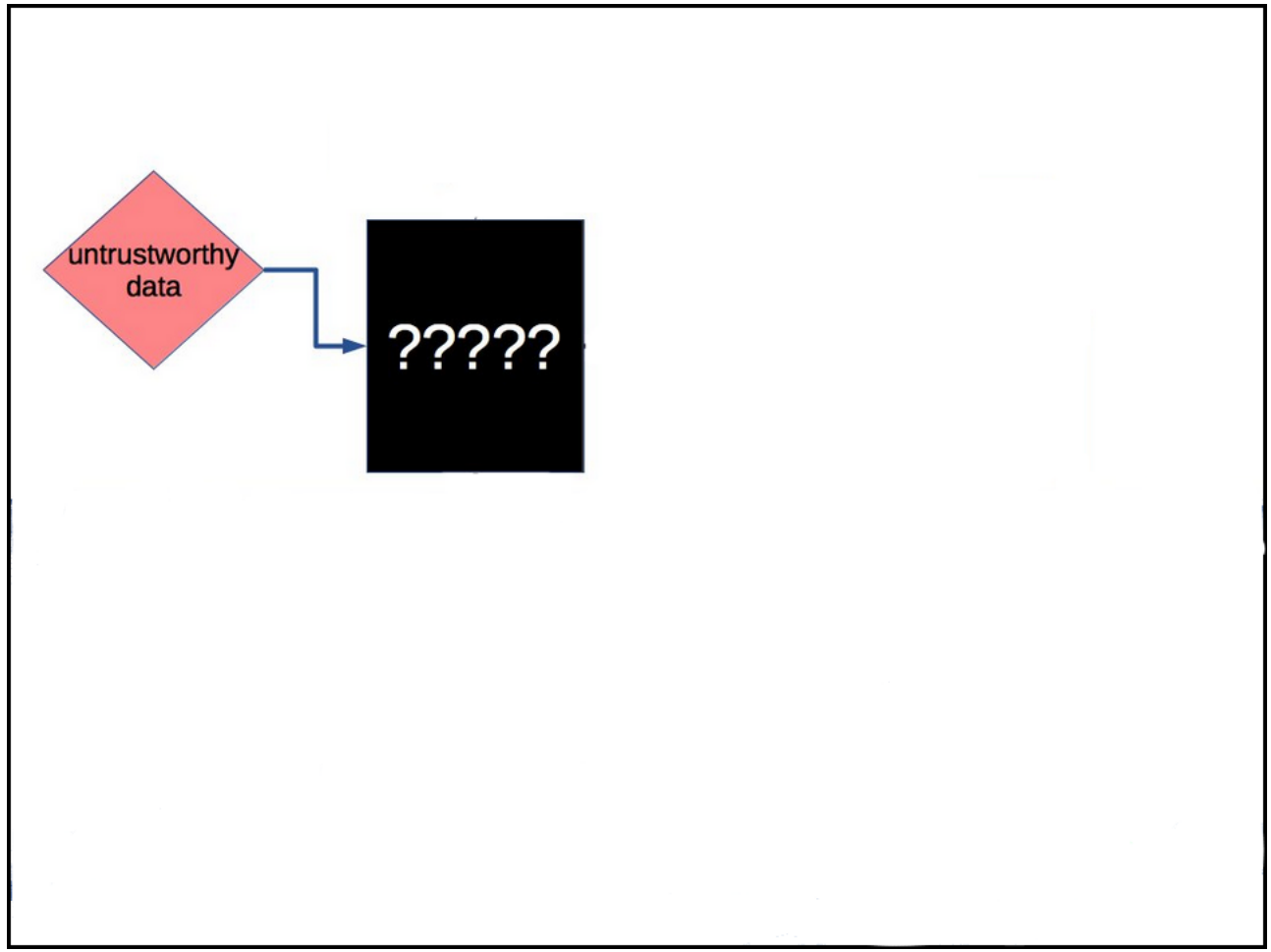
-- Ke Jie

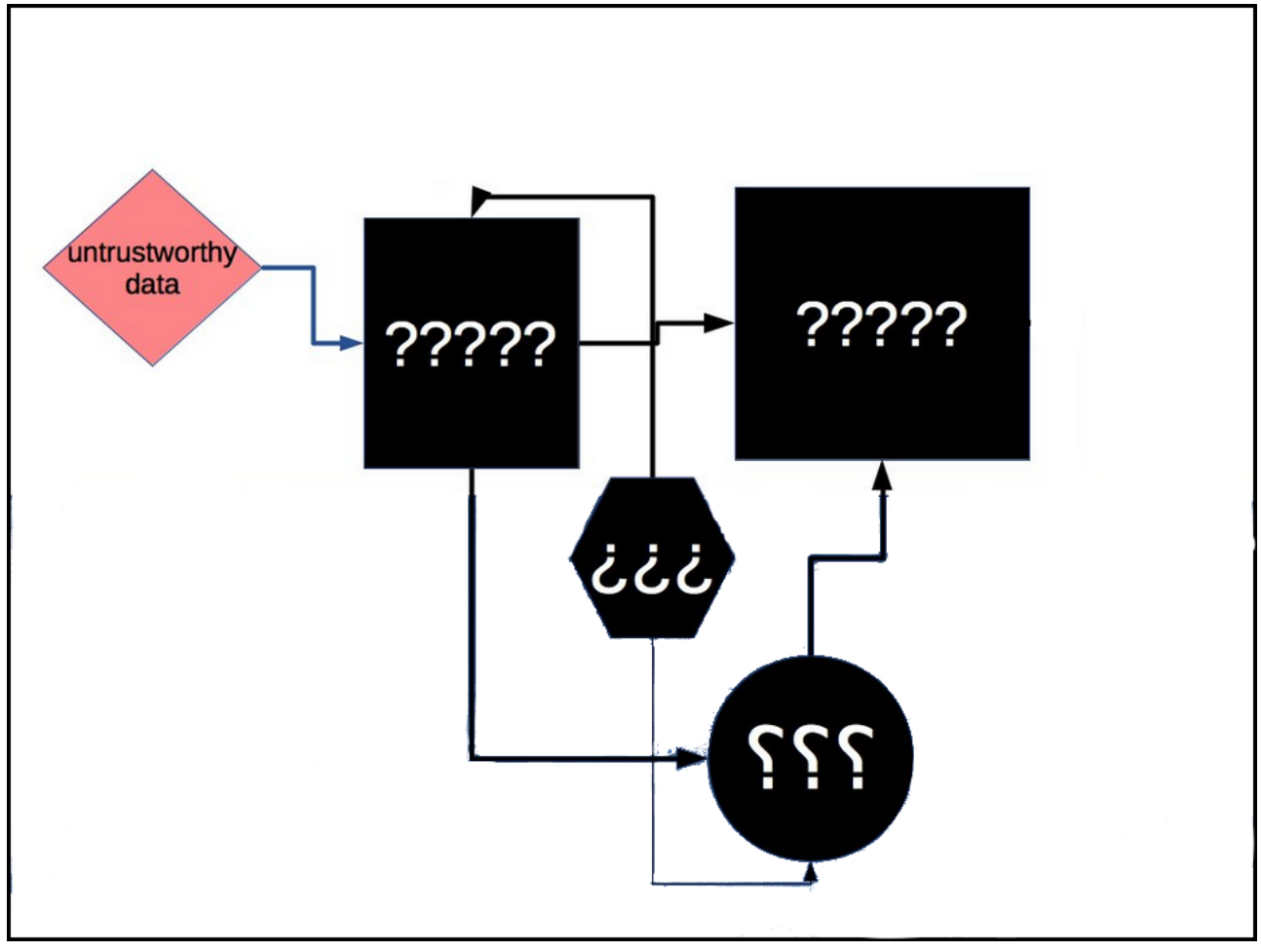
Die Fehler von Maschinellem Lernen

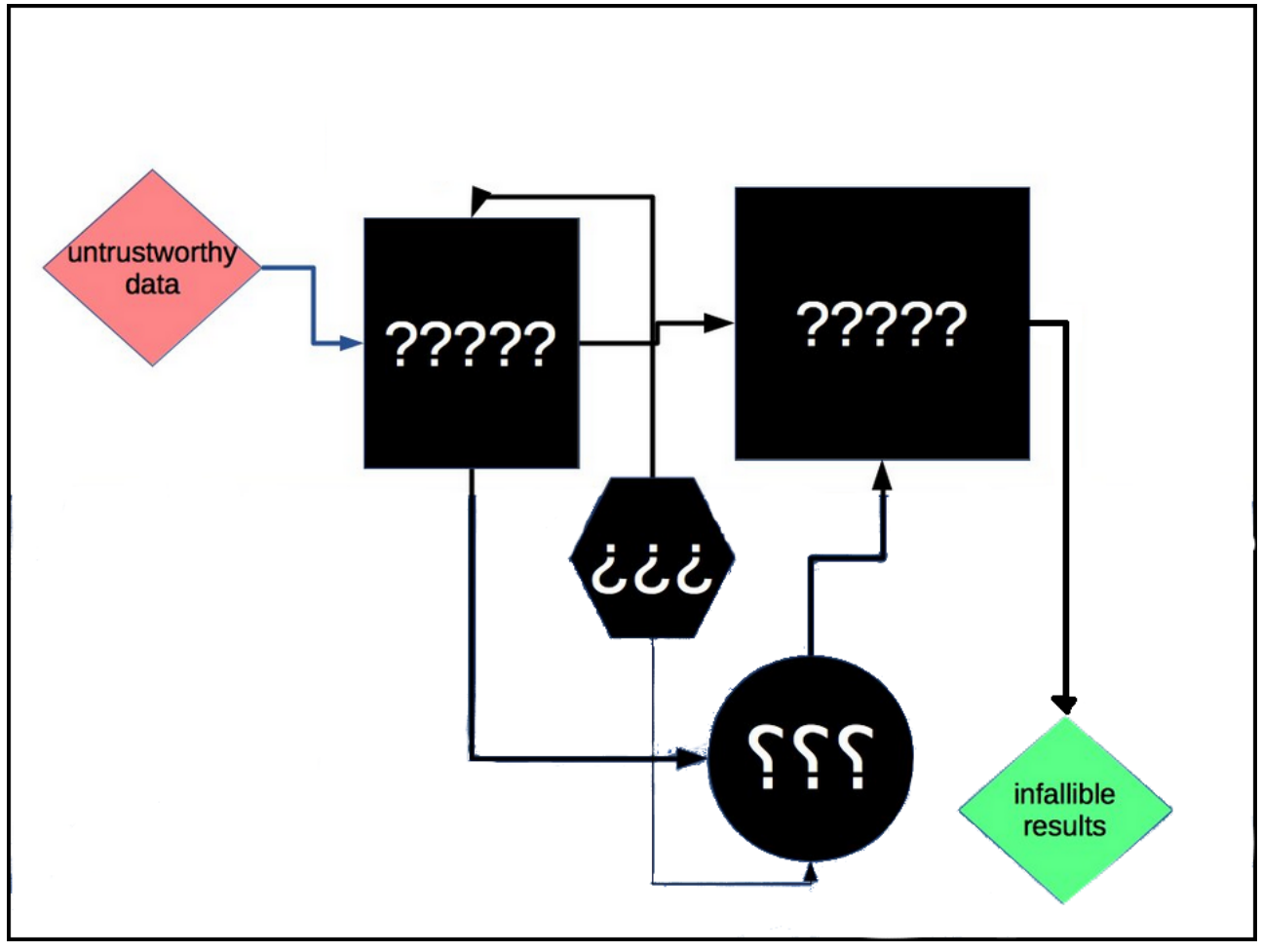

"What is the cost of a train ticket from Sydney to Brisbane"
tap to edit
The answer is about 28.8 quadrillion kilometer US dollars squared.
Input interpretation
The Train (movie) | production budget
The Train (movie) | total US box office receipts
distance | from | Sydney, New South Wales
distance | to | Brisbane, Queensland
Result
28.8488 quadrillion km ^{2} (kilometer US dollars squared)
WolframAlpha
?

ffoo


untrustworthy
data


untrustworthy
data
?????

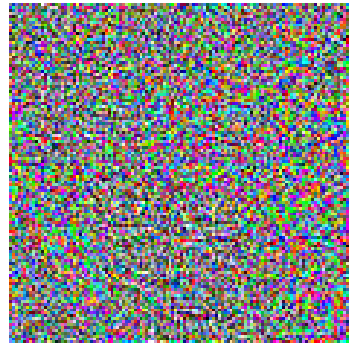

untrustworthy data
?????
?????
¿¿¿
¿¿¿


untrustworthy data
?????
?????
¿¿¿
¿¿¿
infallible results

$\boldsymbol{x}$

"panda"
57.7% confidence

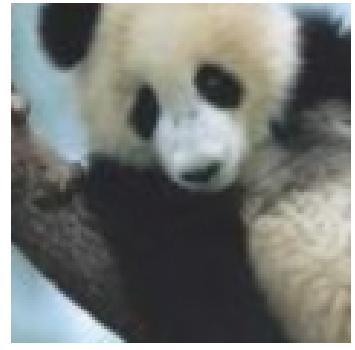
$+ .007 \times$



$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

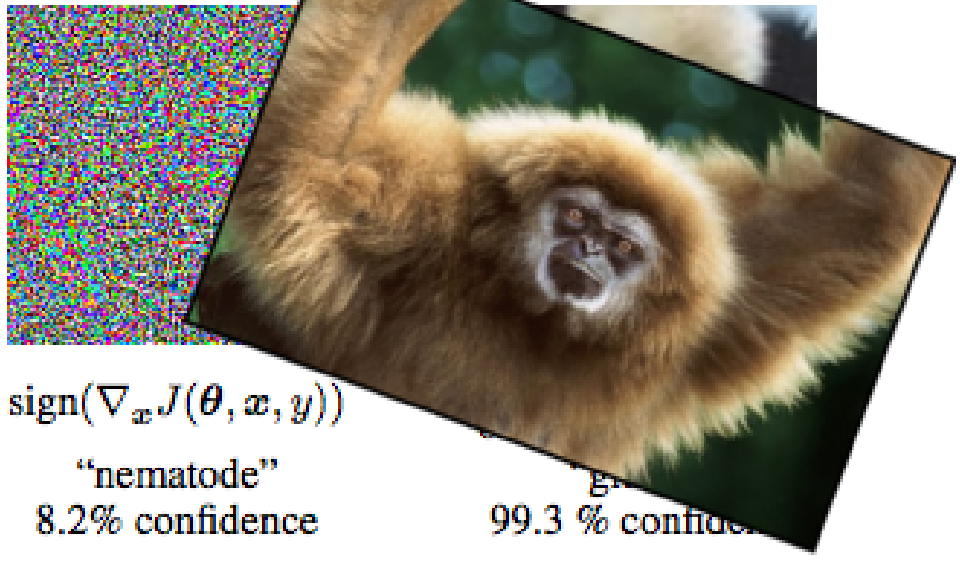
"nematode"
8.2% confidence

$=$



$\boldsymbol{x} +$
$\epsilon \text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"gibbon"
99.3 % confidence

$+ .007 \times$



$\boldsymbol{x}$

"panda"
57.7% confidence

$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$

"nematode"
8.2% confidence

99.3 % confid

Wat lernt misch datt?

Maschinelles Lernen liefert oft
nur *Ergebnisse*, keine *Begründungen*.

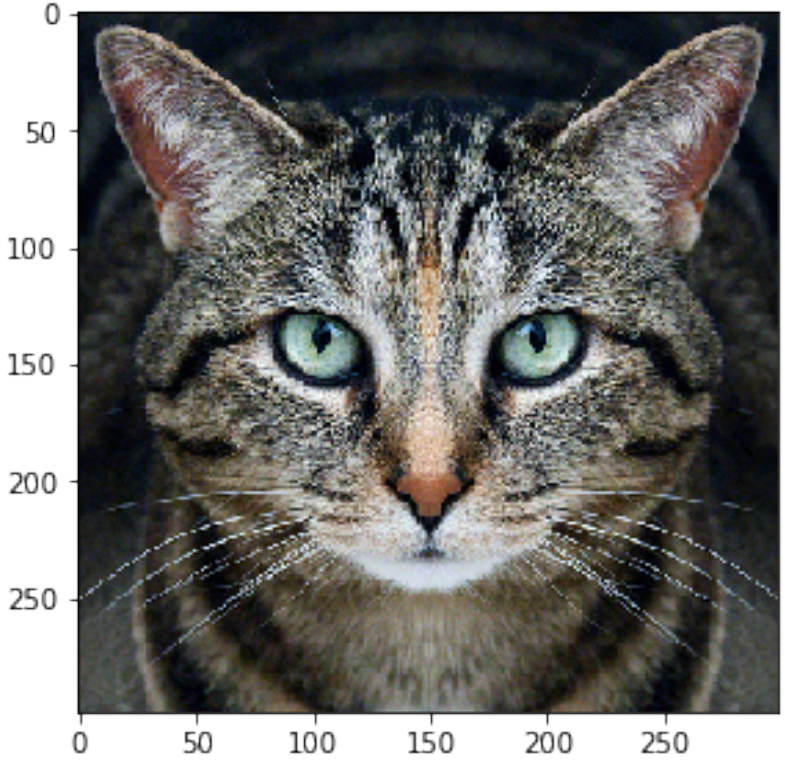
Wat lernt misch datt?

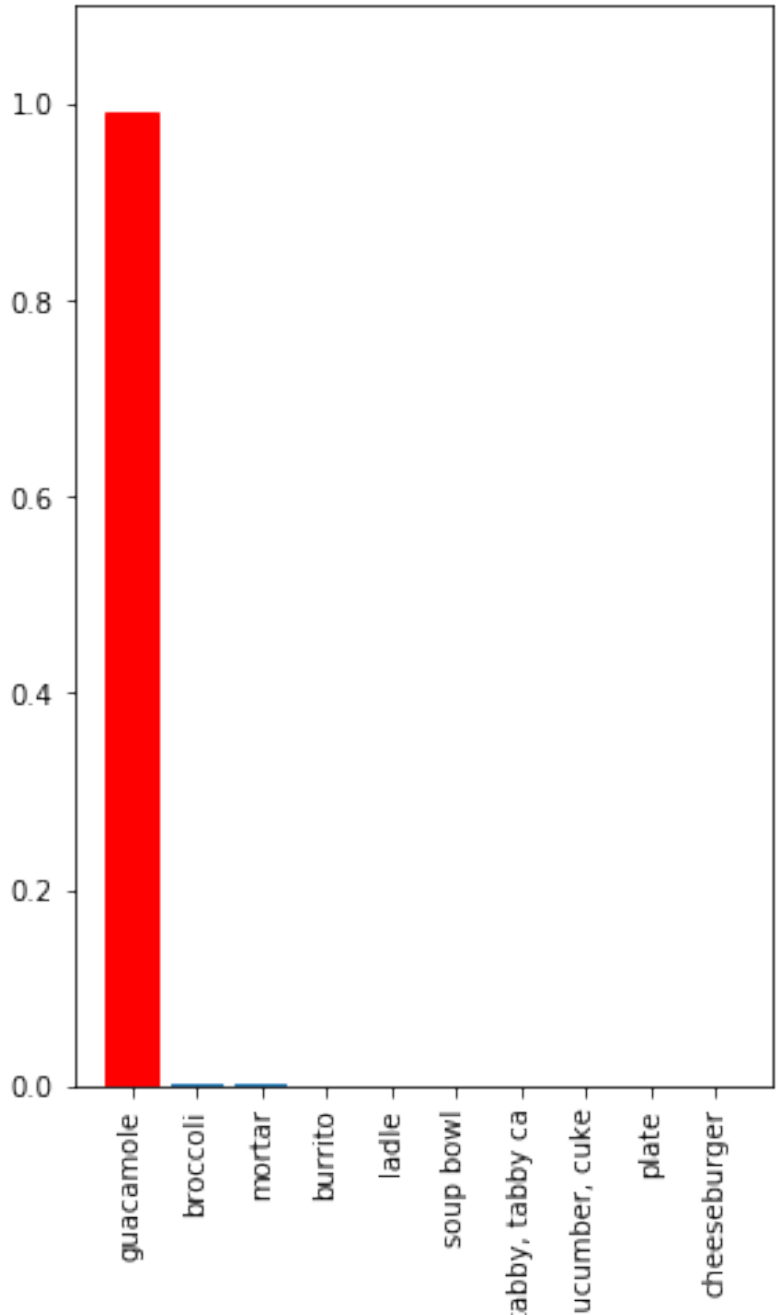
Maschinelles Lernen liefert oft
nur *Ergebnisse*, keine *Begründungen*.

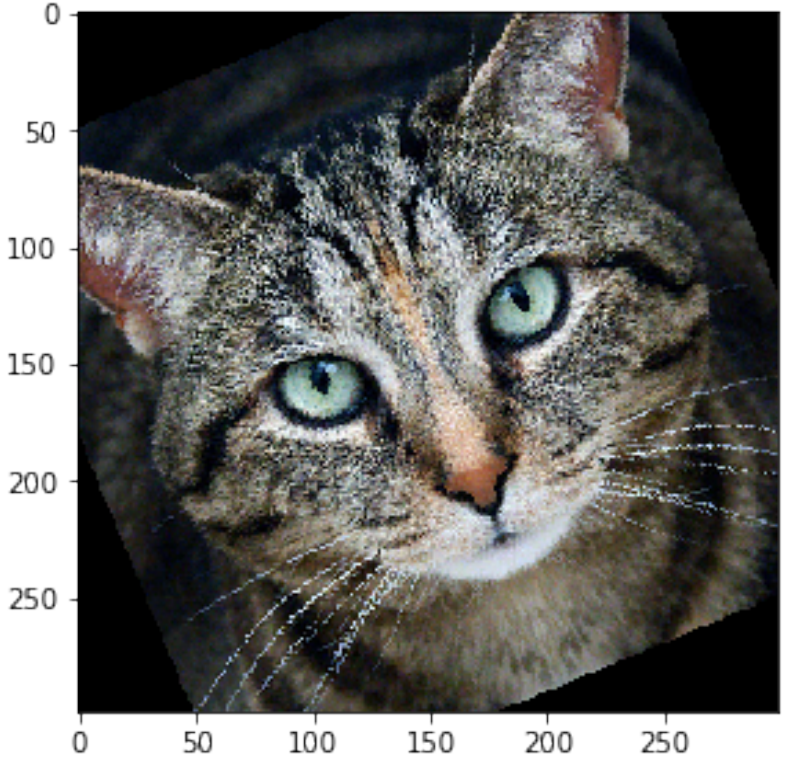
Außerdem ist das Ergebnis höchstens
so allgemein wie die Trainingsdaten.

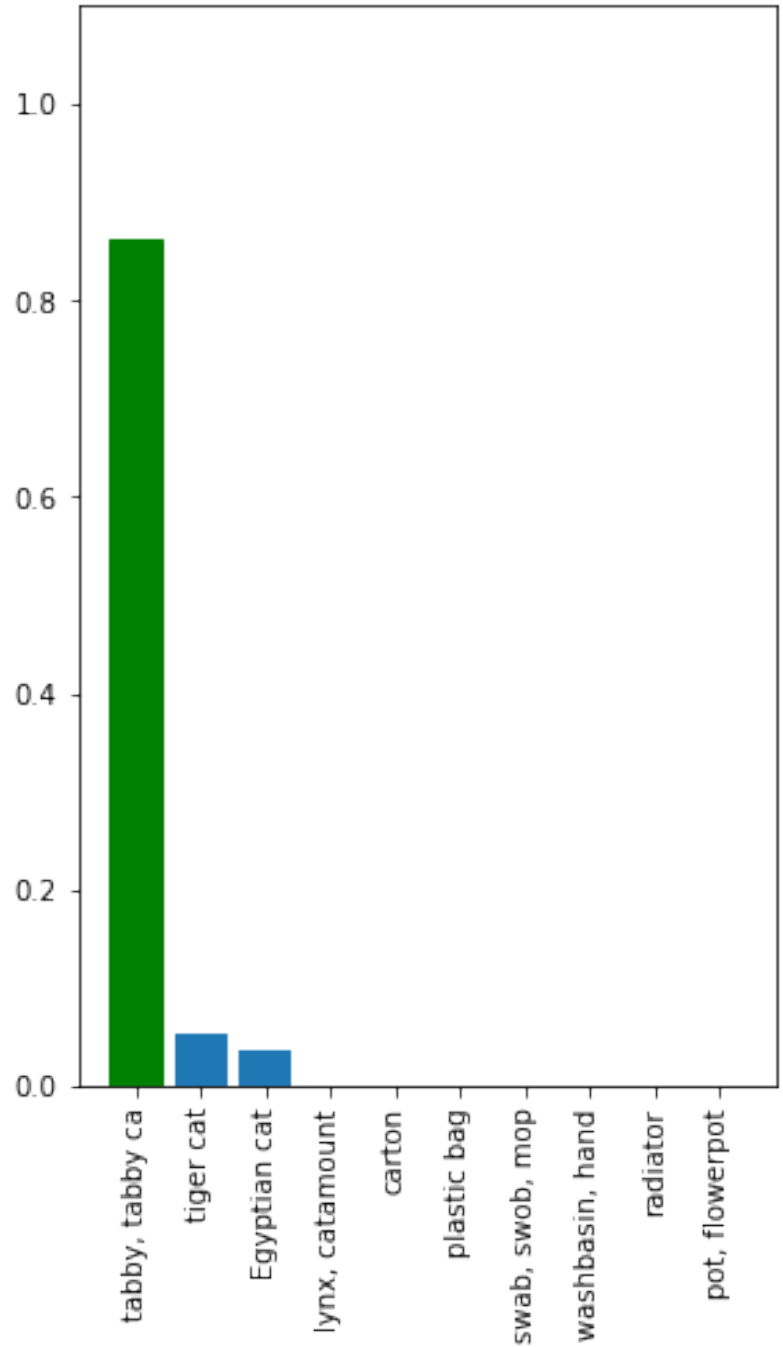
„Adversarial Objects“
(Feindliche Objekte)

(*Subs., plural*) Objekte, die für das menschliche Auge herkömmlich erscheinen, aber für den Computer radikal anders aussehen.


0
50
100
150
200
250
0
50
100
150
200
250


1.0
0.8
0.6
0.4
0.2
0.0
guacamole
broccoli
mortar
burrito
ladle
soup bowl
tabby, tabby ca
cucumber, cuke
plate
cheeseburger


0
50
100
150
200
250
0
50
100
150
200
250


1.0
0.8
0.6
0.4
0.2
0.0
tabby, tabby ca
tiger cat
Egyptian cat
lynx, catamount
carton
plastic bag
swab, swob, mop
washbasin, hand
radiator
pot, flowerpot

Feindliche 3D-gedruckte Schildkröte:

Ein paar offensichtliche Probleme:

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden
- Lars
- ...

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden
- Lars
- ...

Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden
- Lars
- ...

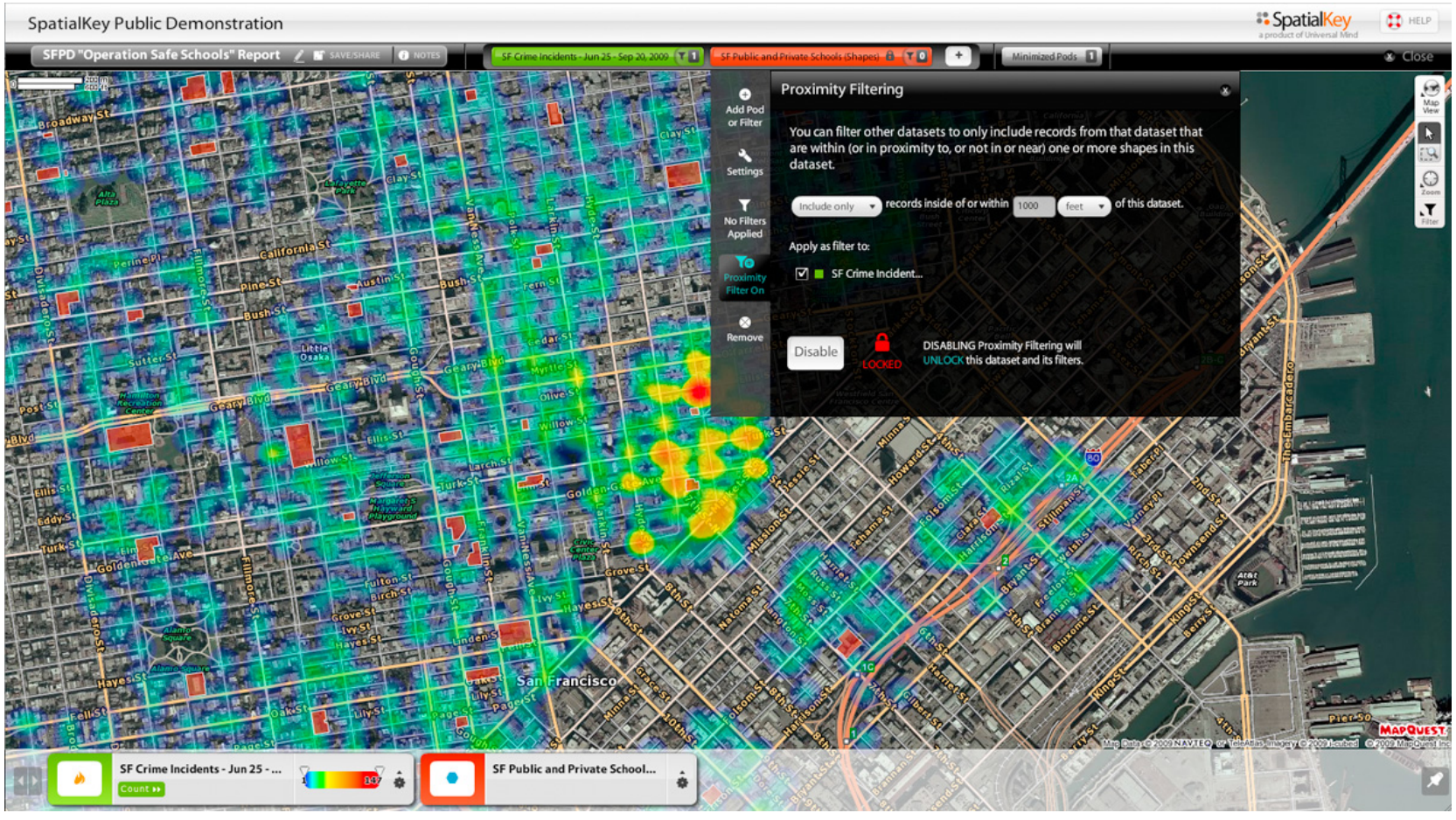
Ein paar offensichtliche Probleme:

- Gratis T-Shirts die am Flughafen als Waffen erkannt werden
- Plakate neben der Autobahn die als Stoppschilder erkannt werden
- LARs
- ...

Wat lernt misch datt?

Maschinelles Lernen ist nicht unfehlbar und darf in kritischen Systemen nie unüberprüft wichtige Entscheidungen treffen.

Fallbeispiel: „Predictive Policing“



Wat lernt misch datt?

Maschinelles Lernen kann (potentiell) Vorurteile und Fehler in der Datengrundlage und im Modell verstärken oder verschlimmern.

Zusammenfassung