HdM Stuttgart
Accessibility Day
17.5.2019

AI Bias

Niklas Janssen & Patrick Brenner

Face-recognition software is perfect – if you're a white man

18

TECHNOLOGY 13 February 2018

https://www.newscientist.com/article/2161028-face-recognition-software-is-perfect-if-youre-a-white-man/

There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks."

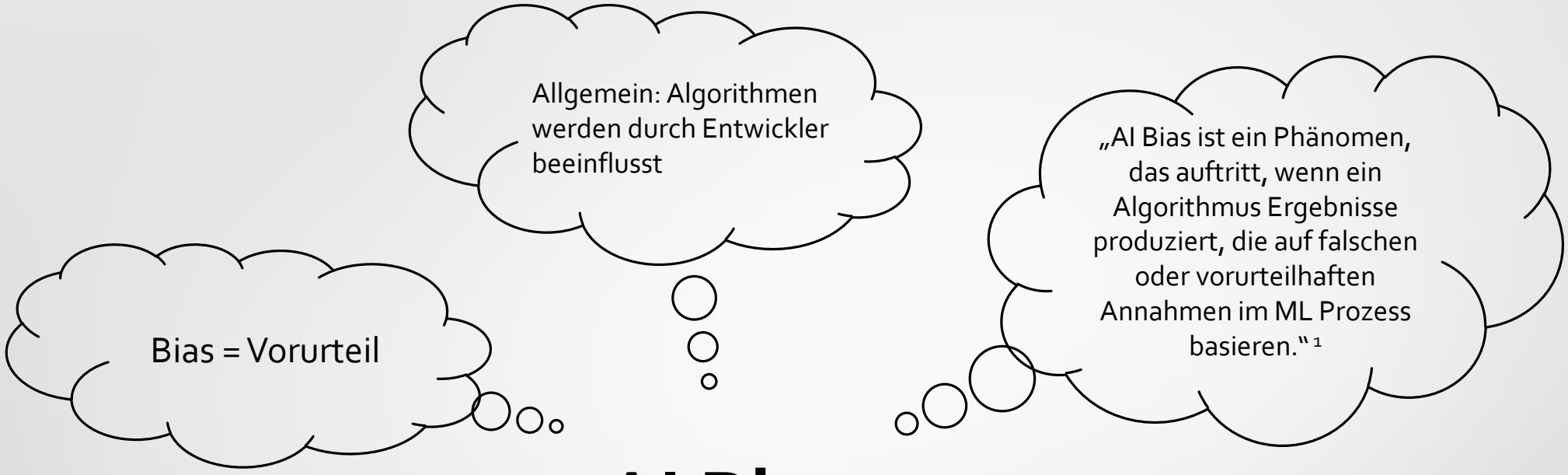
Quelle: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women

BI Isobel Asher Hamilton, Business Insider

10.10.2018, 11:47

Quelle: https://www.businessinsider.de/amazon-built-ai-to-hire-people-discriminated-against-women-2018-10?r=US&IR=T



AI Bias

Diskriminierung, Benachteiligung, Ausschluss von Personengruppen

1: https://searchenterpriseai.techtarget.com/definition/machine-learning-bias-algorithm-bias-or-AI-bias

Wieso tritt AI Bias auf?

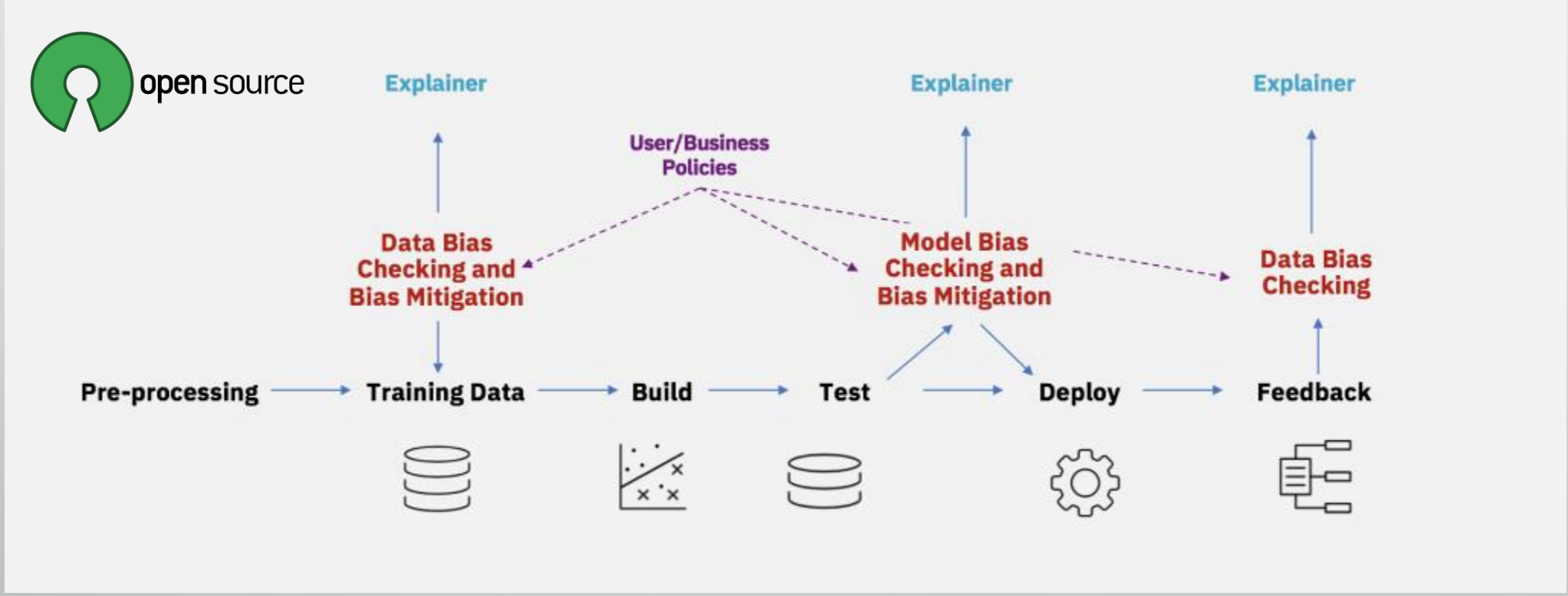
AI Bias kann an verschiedenen Stellen in einen Algorithmus kommen:

- Wahl der berücksichtigten Attribute
- Bewertung der Ergebnisse
- Definition der Zielsetzung
- Auswahl der Trainingsdaten

Wie lässt sich AI Bias verhindern?

- Überprüfen ob Trainingsdaten repräsentativ für Nutzer sind
- Vorsicht bei historischen Datensätzen
- „Human in the loop“
- Bias manuell ausgleichen
- IBM AI Fairness 360

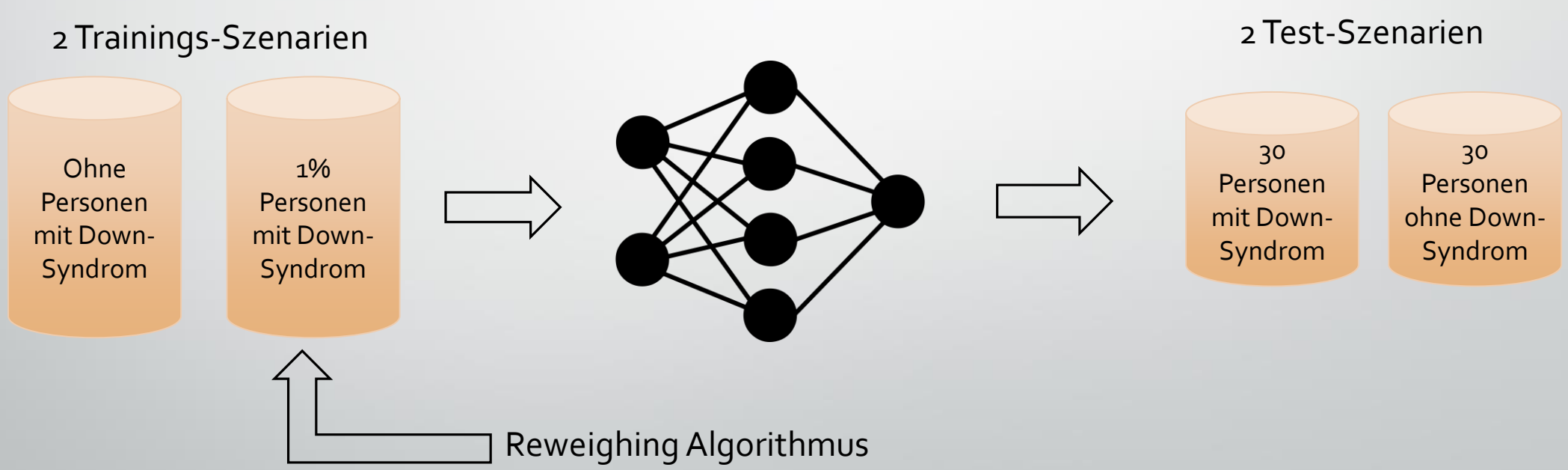
IBM Fairness 360 Toolkit



Quelle: https://www.ibm.com/blogs/research/2018/09/ai-fairness-360/

Projekt: Impact of the IBM AI Fairness 360 toolkit for outliers

https://github.com/pb055/IBMFairness360_DownSyndrom_Test



Projektergebnisse

- Personen mit Down Syndrom, mit 1% Down Syndrom Trainingsbilder

	Without reweighing:	With reweighing:
Correctly predicted:	22	25
Falsely predicted:	8	5
Gender classification accuracy:	73,33%	83,33%
Balanced classification accuracy:	73,33%	83,33%
True positive rate:	86,67%	80,00%
True negative rate:	60,00%	86,67%
Female for male:	2/15	3/15
Male for female:	6/15	2/15

➔ Algorithmus führt zur Verbesserung