Mehr Beispiele für Bias

Cathy O'Neil
https://weaponsofmathdestructionbook.com/

Erkennen von Menschen auf Fotos
https://gizmodo.com/camera-misses-the-mark-on-racial-sensitivity-5256650
https://www.wired.com/2009/12/hp-notebooks-racist/

Gesichtserkennung
https://www.bbc.com/news/technology-50865437

Praxisbeispiel für ein Recidivism Modell
Can you make AI fairer than a judge? Play our courtroom algorithm game
https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/?utm_source=newsletters&utm_medium=email&utm_campaign=the_download.unpaid.engagement

COMPAS
https://en.wikipedia.org/wiki/COMPAS_(software)

Werbung zielt auf psychisch kranke Menschen ab
Vortrag zum Thema Ethik in Marketing und Personalisierter Werbung
https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads#t-704262

Predictive Policing
https://www.themarshallproject.org/2016/02/21/what-you-need-to-know-about-predictive-policing#.5cRMvwYdq

Beispiel für fatales Modell aus dem amerikanischen Gesundheitssystem aus *The Verge*
https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy

Googles Fotoerkennung
https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/

Forschungsarbeiten und Vorträge

Caruana et al.
Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission
http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf

SHAP Paper
Scott Lundberg, Su-In Lee
A Unified Approach to Interpreting Model Predictions
https://arxiv.org/abs/1705.07874

LIME Paper
Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin
"Why Should I Trust You?": Explaining the Predictions of Any Classifier
https://arxiv.org/abs/1602.04938

"Why should I trust you?" - Vortrag zum Thema LIME
https://www.youtube.com/watch?v=KP7-JtFML04

The Great AI Debate - NIPS2017
The first ever debate at a Neural Information Processing Systems conference.
Position: Interpretability is necessary for machine learning
https://www.youtube.com/watch?v=93Xv8vJ2acI

Reduzierung von Bias in NLP Modellen
Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, Kai-Wei Chang
Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints
https://arxiv.org/abs/1707.09457

Anwendungsbeispiele für die Interpretation von Modellen

Kaggle
https://www.kaggle.com/learn/machine-learning-explainability

Beispiel mit R im Blog von Kasia Kulma
https://kkulma.github.io/2017-11-07-automated_machine_learning_in_cancer_detection/

Causation or Correlation?

Keynote: Judea Pearl - The New Science of Cause and Effect
https://www.youtube.com/watch?v=ZaPV1OSEpHw

Bias in Forschung und Industrie

https://www.theguardian.com/lifeandstyle/2019/feb/23/truth-world-built-for-men-car-crashes