LEADERSHIP OF RESPONSIBLE AI:

Artificial Intelligence Bias

Elizabeth M. Adams

LEADERSHIP OF RESPONSIBLE AI:

Intro
AI BIAS

AI technologies are increasingly becoming a part of daily life. Although AI has the potential to be fundamentally objective and without bias, this does not exclude AI bias from existing in the input data that is used to train an AI system or from being encoded in AI algorithms and may even be reinforced in AI outputs (Haenlein & Kaplan, 2019).

IS research has mainly concentrated on the positive effects of IT deployments; one cannot easily find a paper on algorithmic computing and social and human issues that does not mention positive expectations in one form or another (Agerfalk et al., 2021).

Intro

AI BIAS CONTINUED

In numerous high-profile situations, AI has also had unintended societal consequences. These reports merely scratch the surface since most issues go unnoticed.

Some scholars believe we must accept the darker side of AI as a normal part of the design and development process. This viewpoint is motivated by several reported detrimental and unforeseen repercussions that surfaced early on when businesses started deploying and employing AI (Neubert & Montaez, 2020).

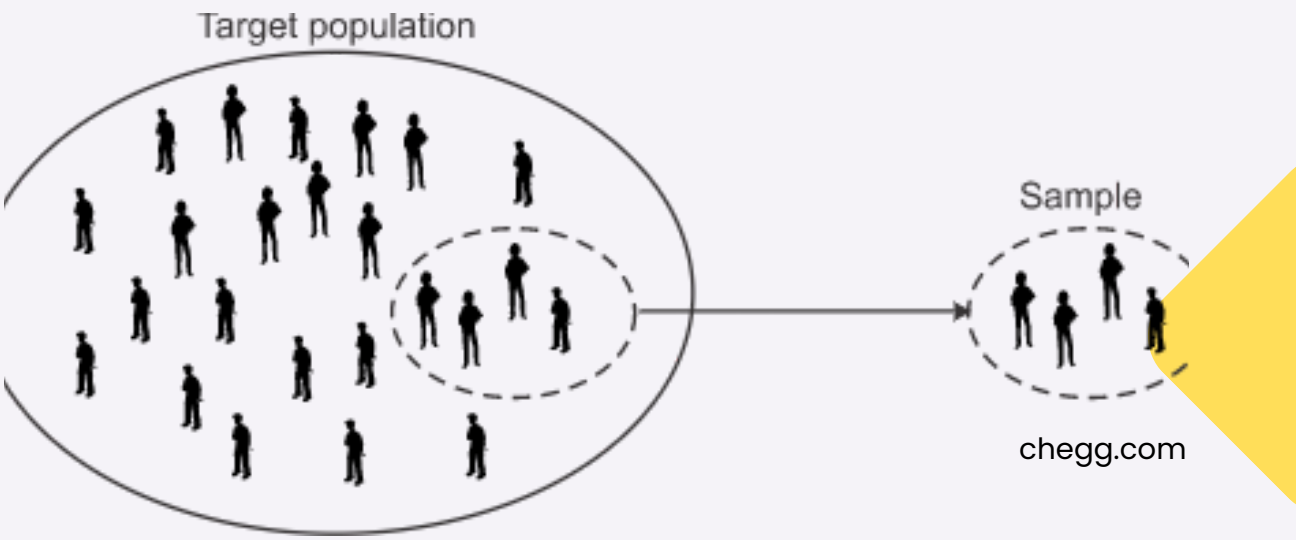
A list of AI bias types is captured on the following pages.

LEADERSHIP OF RESPONSIBLE AI:

SAMPLING BIAS

Produces models relying on training data that is not representative of future cases.

Mirbabaie et al. (2022) and Zhou et al. (2022)



LEADERSHIP OF RESPONSIBLE AI:

CONFIRMATION BIAS

Leads to machine learning searches that reinforce biases

Mirbabaie et al. (2022) and Zhou et al. (2022)

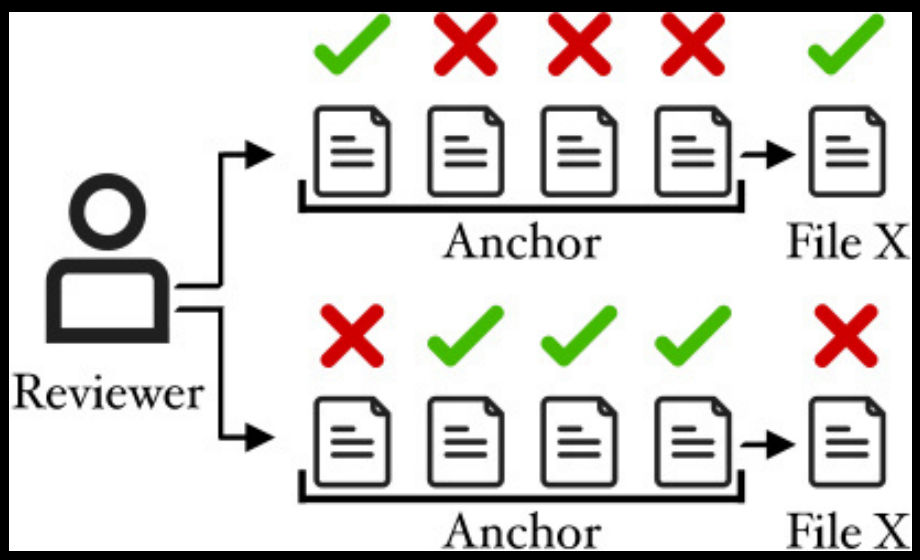
Confirmation bias, is our propensity to believe information that supports our preexisting opinions and ignore information that contradicts them.

LEADERSHIP OF RESPONSIBLE AI:

ANCHORING BIAS

Leads to incorrect assumptions about initial information provided by AI

Mirbabaie et al. (2022) and Zhou et al. (2022)

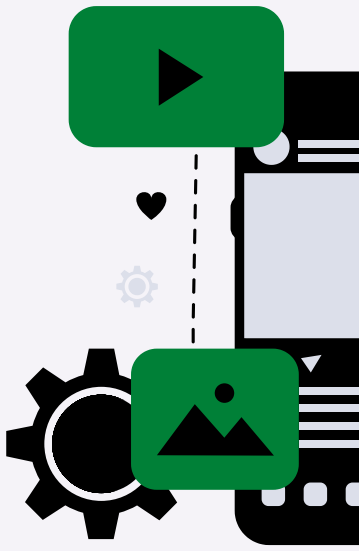


https://dl.acm.org/doi/fullHtml/10.1145/3491102.3517443

As a subset of automation bias, which occurs when people place an inappropriate amount of trust in automation. Some people do not measure their trust in AI based on its reliability.

LEADERSHIP OF RESPONSIBLE AI:

COMPUTER BIAS



A framework by Friedman and Nissenbaum (1996) with defined criteria for reliability, accuracy and efficiency of which computer systems should be judged.

Mirbabaie et al. (2022) and Zhou et al. (2022)

Three categories of bias in **computer systems** have been developed: **preexisting**, **technical**, and **emergent**.

Preexisting bias has its roots in social institutions, practices, and attitudes.

Technical bias arises from technical constraints of considerations.

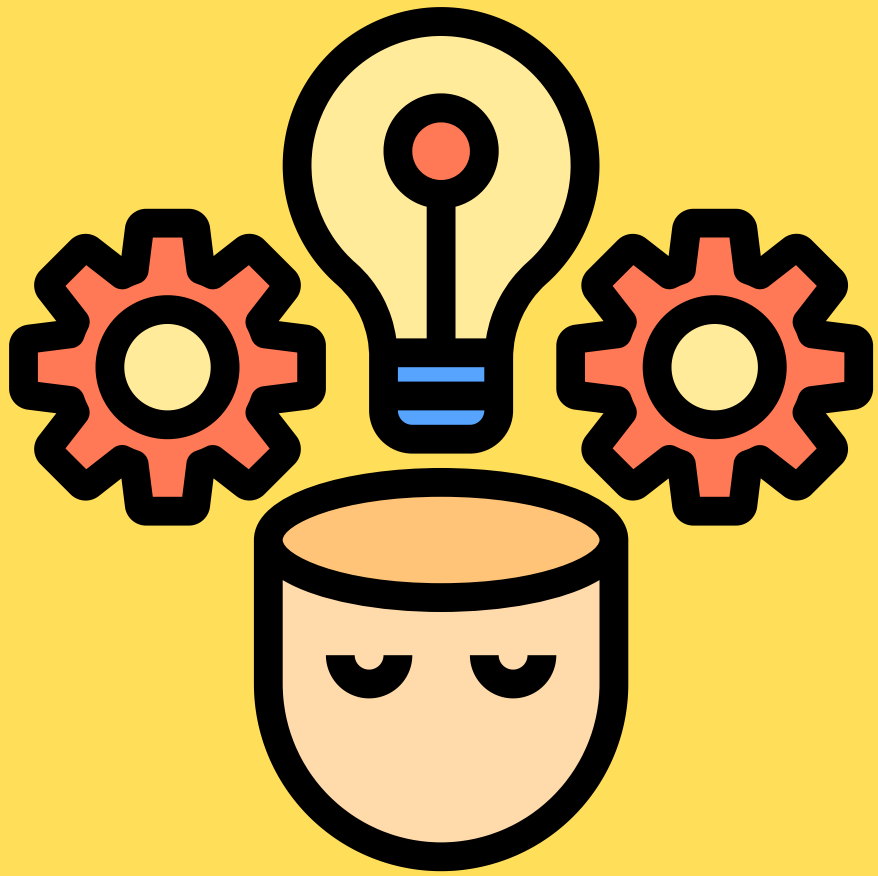
Emergent bias arises in a context of use.

LEADERSHIP OF RESPONSIBLE AI:

PERFORMANCE BIAS

Examines performance distortion in predictions by AI

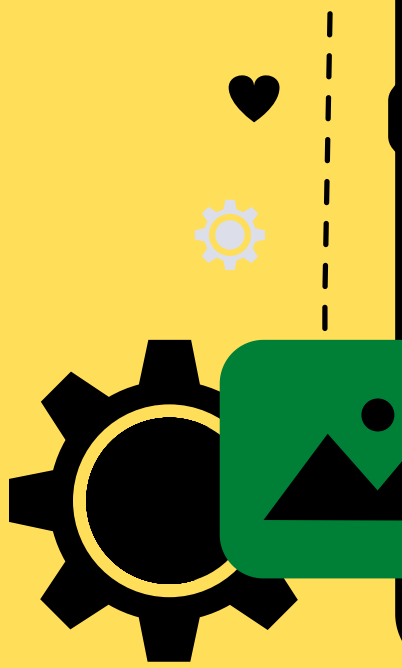
Mirbabaie et al. (2022) and Zhou et al. (2022)



Perceptions of predictive ability, generalizability, and performance consistency across data subsets can be inflated by performance bias.

Abassai et al., 2018

LEADERSHIP OF RESPONSIBLE AI:

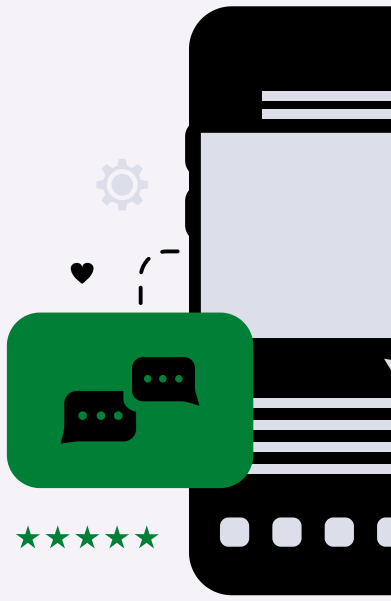


DATA BIAS

a. Bias in historical data leading to discrimination
b. Bias in data collection mechanisms that lack representativeness
c. Bias in alternate sources of data that predict outcomes
d. Bias in unobservable outcomes making it difficult to measure discrimination
e. Bias in unstructured data and feature engineering to include but not limited to texts, audio, images analyzed through AI/ML

Mirbabaie et al. (2022) and Zhou et al. (2022)

LEADERSHIP OF RESPONSIBLE AI:



ALGORITHMIC BIAS (VIA MACHINE LEARNING)

a. Modern ML algorithms using millions of observations as predictors
b. Overfitting of training data with ML algorithms remembering patterns leading to concerns of privacy
c. Optimization of ML with a focus on maximizing predictive performance while ignoring fairness
d. Poor optimization with Data Bias causing harm to protected groups especially when such groups are under-represented in the training data.
e. Lack of interpretability of complex ML algorithms unable to identify the presence or reasons for discrimination

Mirbabaie et al. (2022) and Zhou et al. (2022)

LEADERSHIP OF RESPONSIBLE AI:

MIRBABAIE ET AL. (2022) PROVIDE A COMPREHENSIVE LIST OF BIASES FOUND IN AI TECHNOLOGIES WHICH INCLUDE SAMPLING BIAS, PERFORMANCE BIAS, CONFIRMATION BIAS AND ANCHORING BIAS.

MIRBABAIE, M., BRENDEL, A. B., & HOFEDITZ, L. (2022). ETHICS AND AI IN INFORMATION SYSTEMS RESEARCH. COMMUNICATIONS OF THE ASSOCIATION FOR INFORMATION SYSTEMS, 50, PP-PP. HTTPS://DOI.ORG/10.17705/1CAIS.05034

LEADERSHIP OF RESPONSIBLE AI:

ZHOU ET AL. (2022) PROVIDE A DETAILED VIEW OF POTENTIAL SOURCES FOR BIAS AND DISCRIMINATION, INCLUDING DATA AND ALGORITHMIC BIAS. THEY UNCOVER HOW BIAS FORMS FROM HISTORICAL DATA AND IS OFTEN SKEWED TOWARDS OR AGAINST PARTICULAR GROUPS.

ZHOU, N., ZHANG, Z., NAIR, V. N., SINGHAL, H., & CHEN, J. (2022). BIAS, FAIRNESS AND ACCOUNTABILITY WITH ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING ALGORITHMS. INTERNATIONAL STATISTICAL REVIEW. HTTPS://DOI.ORG/10.1111/INSR.12492

LEADERSHIP OF RESPONSIBLE AI:

Elizabeth M. Adams

AI BIAS TYPES

DID YOU FIND THIS HELPFUL?

If you found this post helpful,
don't forget to like, comment,
share, and save it.