

# Is the Dataset Biased?



# Responsible AI: Human-Centered Design



## Course 1

*Fundamentals of TinyML*

- What am I building?
- Who am I building this for?
- What are the consequences for the user if it *fails*?

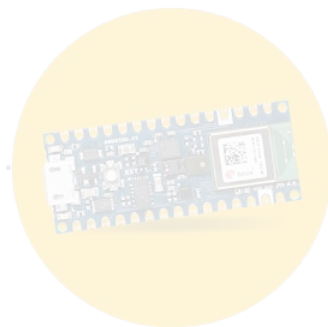
## Course 2

*Applications of TinyML*

- What data will be collected to train the model?
- **Is the dataset biased?**
- How can we ensure the model is fair?

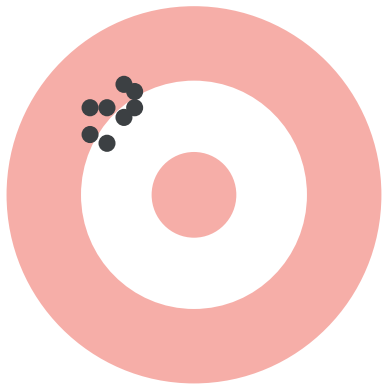
## Course 3

*Deploying TinyML*



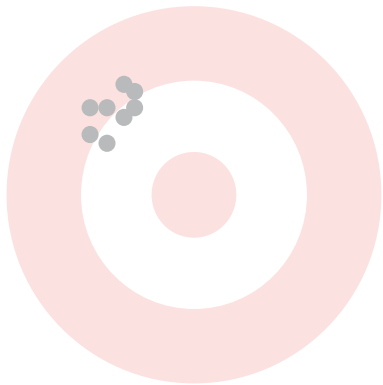
# What is bias?

**Bias** is a deviation in a ***predictable***  
(i.e., not random) direction

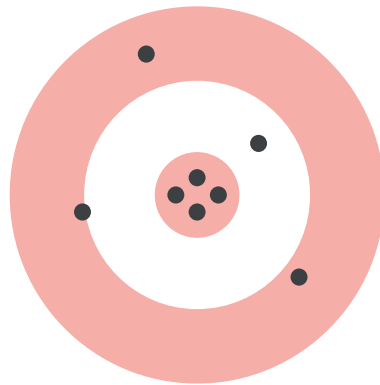


# What is bias?

**Bias** is a deviation in a *predictable*  
(i.e., not random) direction



**Not all errors**  
**are attributed to bias**



# What does **bias** mean for **ML**?

Bias in machine learning = **systematic** errors that lead to inaccurate results

- Biased **datasets** (faulty, poor, or incomplete data) can lead to a distribution **mismatch** between the **dataset** and **reality**
- This may lead to **inaccurate results**, or worse, **discriminatory** or **unfair results**

# The “garbage in, garbage out” problem



# ***Bias:*** Defining the Target Variable

- **How** should you define a “**good**” employee for a hiring algorithm?
- **Subjective** process: “**good**” must be defined in ways that correspond to **measurable** outcomes

|                      |                              |
|----------------------|------------------------------|
| Number of sick days  | Number of times late to work |
| Length of employment | Number of sales per quarter  |

# ***Bias:*** Defining the Target Variable

- **How** should you define a “**good**” employee for a hiring algorithm?
- **Subjective** process: “**good**” must be defined in ways that correspond to **measurable** outcomes



Number of  
sick days

Number of  
times late to  
work

Length of  
employment

Number of  
sales per  
quarter



# ***Bias:*** Defining the Target Variable

Using **biometric** sensors for a health wearable device, how should you define “***healthy***”?

- Heart rate
- Blood pressure
- Number of steps



# ***Bias:*** Labeling the Data

Labels applied to the training data must serve as **ground truth**



Horse



Human



Human

**ERROR**



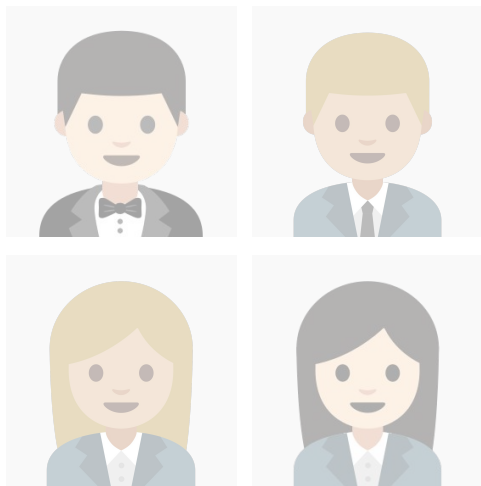
# ***Bias:*** Sampling the Data

## Training Data

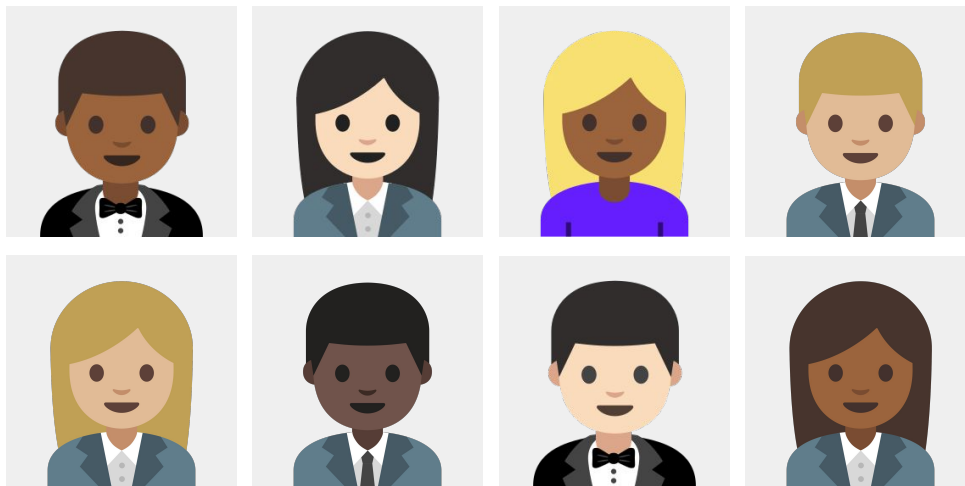


# ***Bias:*** Sampling the Data

Training Data

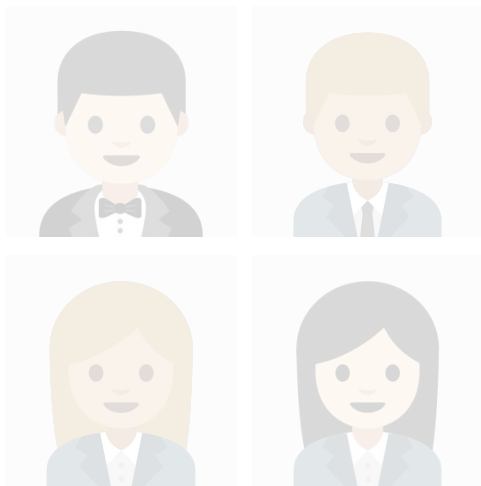


Reality

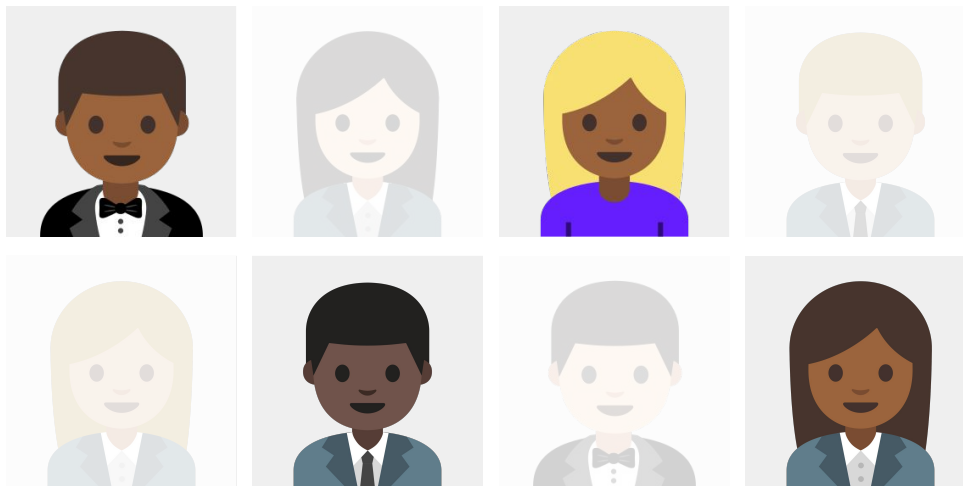


# ***Bias: Sampling the Data***

Training Data

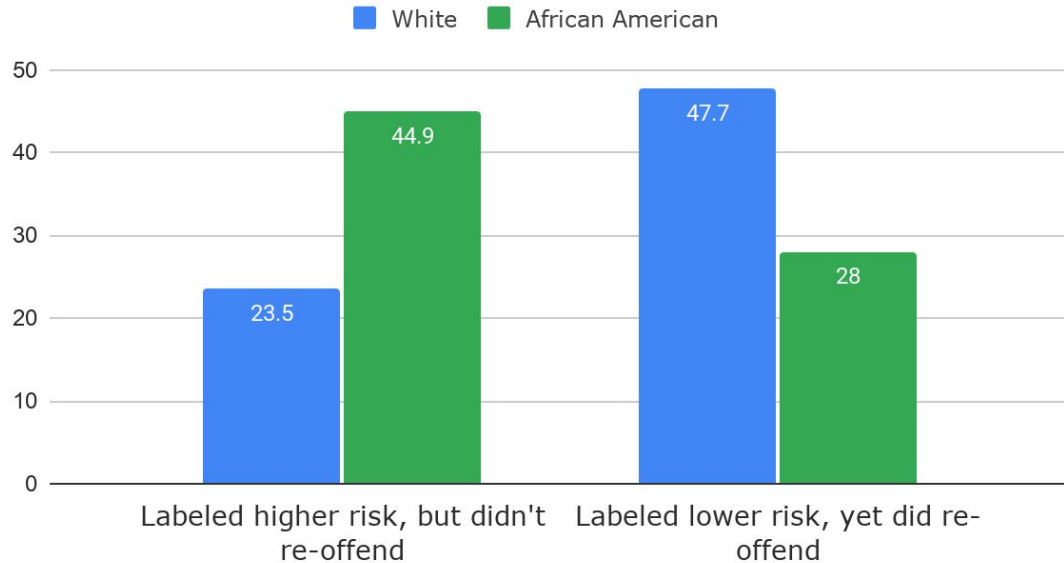


Reality



# Bias: Prejudice Reflected in Data

**COMPAS Risk Assessment %**



**Northpointe's COMPAS**  
Recidivism Prediction Tool

# ***Bias: Prejudice Reflected in Data***



**Dataset:** 65% of people cooking are *women*

**Algorithm predicts:** 85% of people cooking are *women*



# ***Bias:*** Measurement **Distortion**



**Hardware matters!**

**Example:** Optic sensors and cameras have problems ***detecting darker skin tones*** due to the way light is reflected

- **Automatic soap dispensers**
- **Activity detection**
- **Facial recognition**

# ***Bias:*** Measurement **Distortion**



How can we *fix* biased data?

# Industry Solutions: Datasheets for Datasets

Questions for dataset creators to reflect on during the key stages of the dataset lifecycle:

- ***Motivation***
- ***Composition***
- ***Collection Process***
- ***Preprocessing/ labeling***
- ***Uses***
- ***Distribution***
- ***Maintenance***

paper authored by

JAMIE MORGENSTERN, Georgia Institute of Technology  
TIMNIT GEBRU, Google  
BRIANA VECCHIONE, Cornell University  
JENNIFER WORTMAN VAUGHAN, Microsoft Research  
HANNA WALLACH, Microsoft Research  
HAL DAUMÉ III, Microsoft Research; University of Maryland  
KATE CRAWFORD, Microsoft Research; AI Now Institute

# Industry Solutions: Data Nutrition Labels

| Metadata    |  |
|-------------|--|
| Filename    | 201612v1-docdollars-product_payments   |
| Format      | csv  |
| Url         | <a href="https://projects.propublica.org/docdollars/">https://projects.propublica.org/docdollars/</a>  |
| Domain      | healthcare   |
| Keywords    | Physicians, drugs, medicine, pharmaceutical, transactions  |
| Type        | tabular  |
| Rows        | 500  |
| Columns     | 18   |
| Missing     | 5.2%   |
| License     | cc   |
| Released    | JAN 2017   |
| Range       | From   |
|             | AUG 2013   |
|             | To   |
|             | DEC 2015   |
| Description | This is the data used in ProPublica's Dollars for Docs news application. It is primarily based on CMS's Open Payments data, but we have added a few features. ProPublica has standardized drug, device and manufacturer names, and made a flattened table (product_payments) that allows for easier aggregating payments associated with each drug/device. In [1], one payment record can be attributed to up to five different drugs or medical devices. This table flattens the payments out so that each drug/device related to each payment gets its own line. |



A standard label that highlights the “**key ingredients**” of a dataset:

- ***Provenance***
- ***Metadata***
- ***Missing units***
- ***Variables***

# Industry Solutions: **Bias Testing Toolkits**

IBM Research Trusted AI



## AI Fairness 360

This extensible open source toolkit can help you examine, report, and mitigate discrimination and bias in machine learning models throughout the AI application lifecycle. We invite you to use and improve it.



# Designer Solutions

- Carefully **research your users in advance**, be aware of potential outliers
- **Ensure your team** of data scientists and data labelers is *diverse*
- Where possible, **combine inputs from multiple sources** to ensure data diversity
- **Create a gold standard** for data labeling
- Seek out **domain experts** to review your data