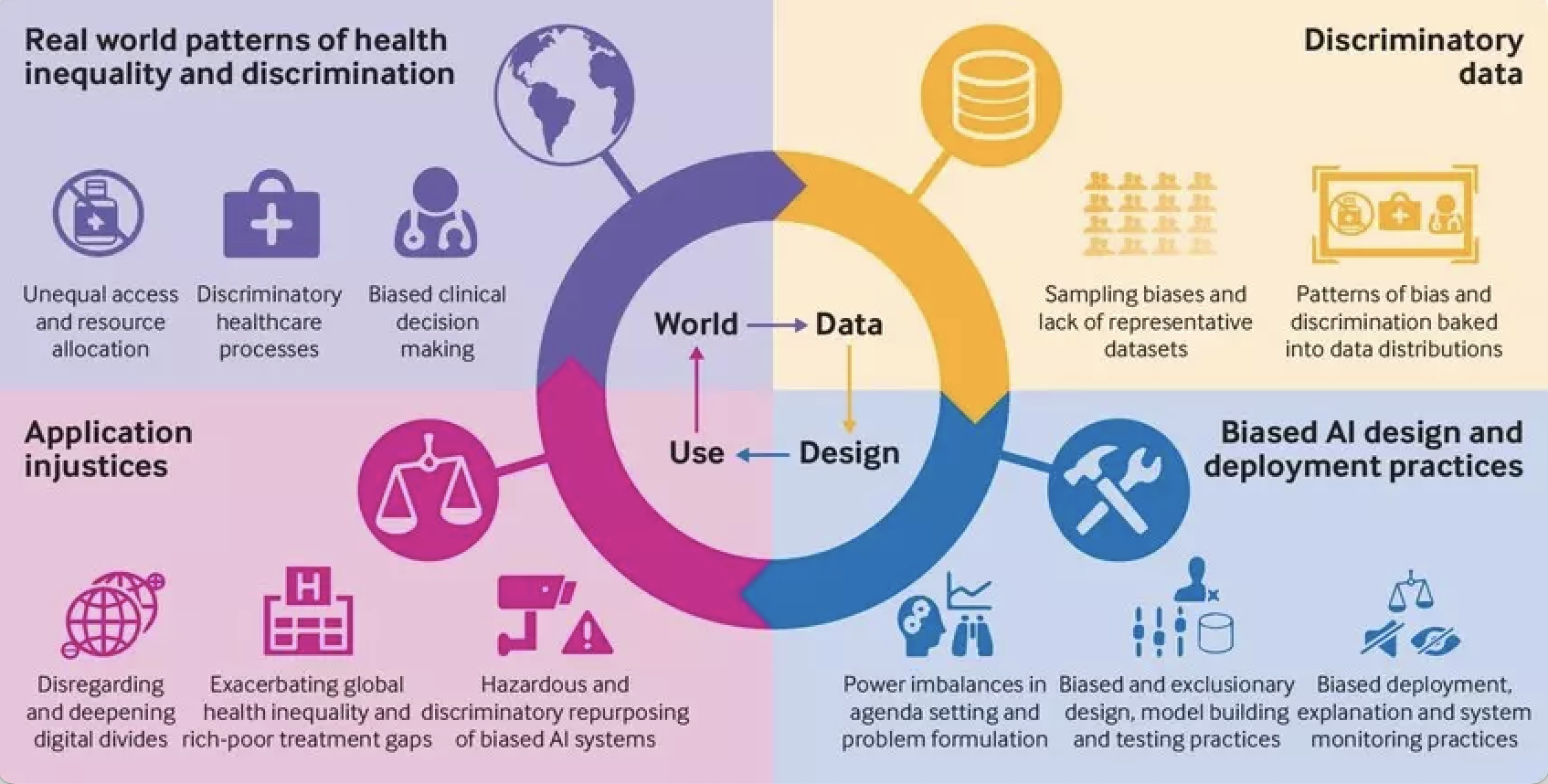
BIAS IN AI TRAINING

Often we hear about implicit bias in the workplace. One implicit bias you may have heard of is favoring familiar sounding names than those from other cultural groups.



*But what about bias in AI training?*

Algorithmic bias describes systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others. Click [here](https://www.forbes.com/sites/forbestechcouncil/2021/02/04/the-role-of-bias-in-artificial-intelligence/?sh=6584f5ee579d) to learn more about the role of bias in AI training.

Review the following [types](https://www.weforum.org/agenda/2021/07/ai-machine-learning-bias-discrimination/) of biases and risks in AI training:

**Implicit bias** is discrimination or prejudice against a person or group that is unconscious to the person with the bias. It is dangerous because the person is unaware of the bias – whether it be on grounds of gender, race, disability, sexuality or class.

**Sampling bias** is a statistical problem where random data selected from the population do not reflect the distribution of the population. The sample data may be skewed towards some subset of the group.

**Temporal bias** is based on our perception of time. We can build a machine-learning model that works well at this time, but fails in the future because we didn't factor in possible future changes when building the model.

**Over-fitting to training data** happens when the AI model can accurately predict values from the training dataset but cannot predict new data accurately. The model adheres too much to the training dataset and does not generalize to a larger population.

**Edge cases and outliers** are data outside the boundaries of the training dataset. Outliers are data points outside the normal distribution of the data. Errors and noise are classified as edge cases: Errors are missing or incorrect values in the dataset; noise is data that negatively impacts on the machine learning process.

Now, we will examine a significant racial bias found in Georgia Institute Technology’s 2019 study on self-driving cars.

Please take 10 minutes to read the following case and let’s dive right into the questions. Click [here](https://www.businessinsider.com/self-driving-cars-worse-at-detecting-dark-skin-study-says-2019-3) to read the case.

Question for Discussion:

1. Does AI use common sense?
2. Where does bias in AI usually come from?
3. What are some ways to prevent bias in AI training?
4. Describe a self-driving scenario, which AI would struggle to respond safely to and why?
5. Referring to the five biases/risks listed earlier, share an example of one in self-driving AI training.

\*How do we circle back to the donkey car?

-

Helpful sources:

<https://www.weforum.org/agenda/2021/07/ai-machine-learning-bias-discrimination/>

<https://drive.google.com/file/d/11WHVbrX3kffUaXkI7-YHo1k4h9TQJ-Pr/view>

<https://www.forbes.com/sites/forbestechcouncil/2021/02/04/the-role-of-bias-in-artificial-intelligence/?sh=6584f5ee579d>