AI Boot Camp

AI Ethics:
An Introduction

Module 15 Day 1

Class Objectives
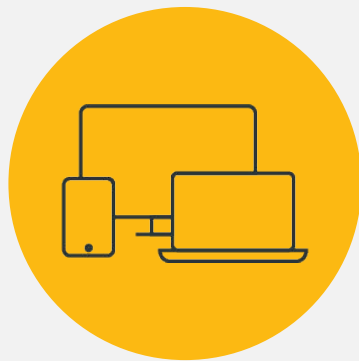
By the end of class, you will be able to:

1. Explain what ethics means.

2. Summarize the major ethical concerns related to AI.

3. Explain why algorithmic bias is such a significant ethical concern.

4. List some common causes of algorithmic bias.

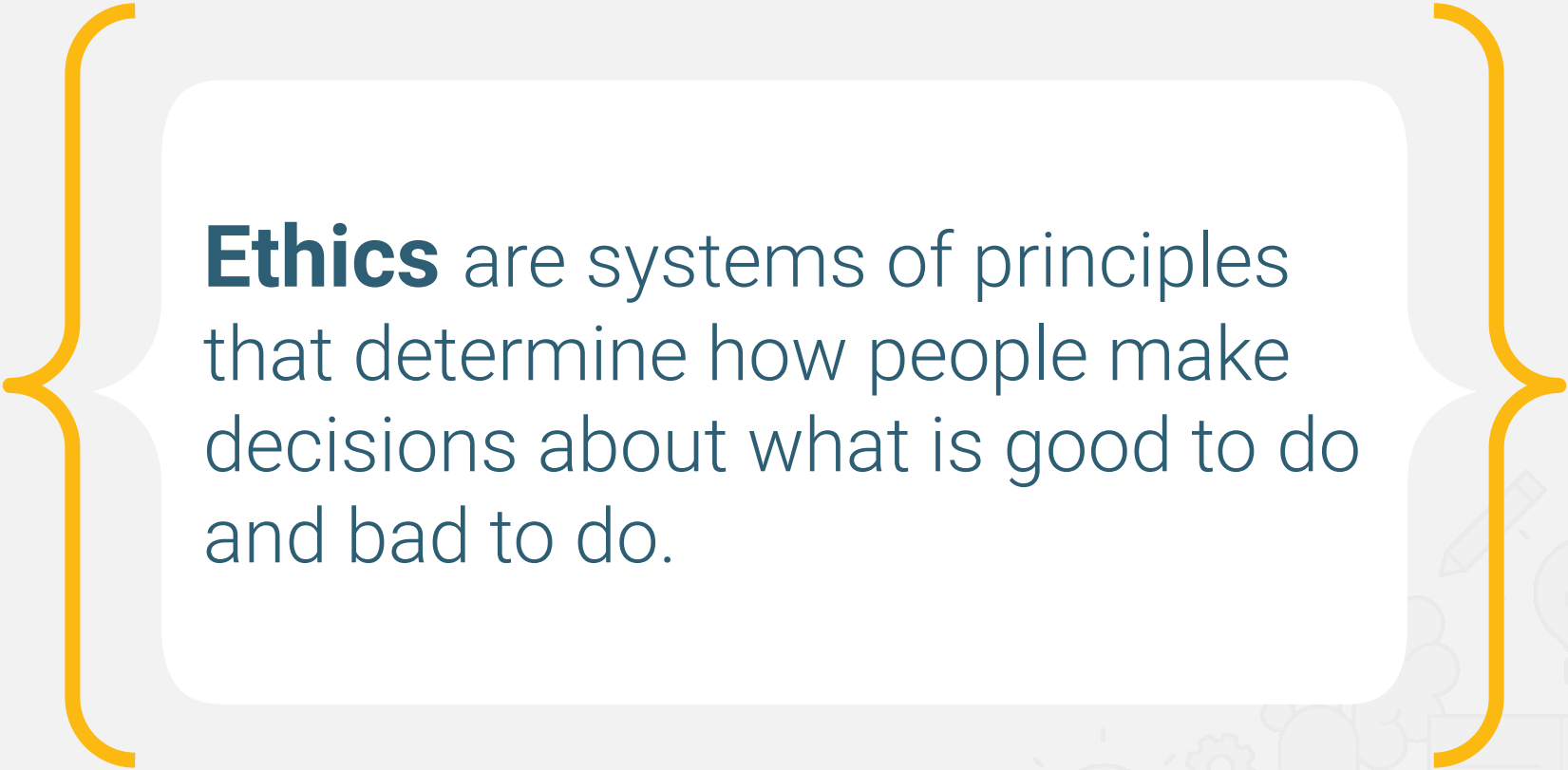5. Analyze a case study in which researchers identified algorithmic bias in a system.
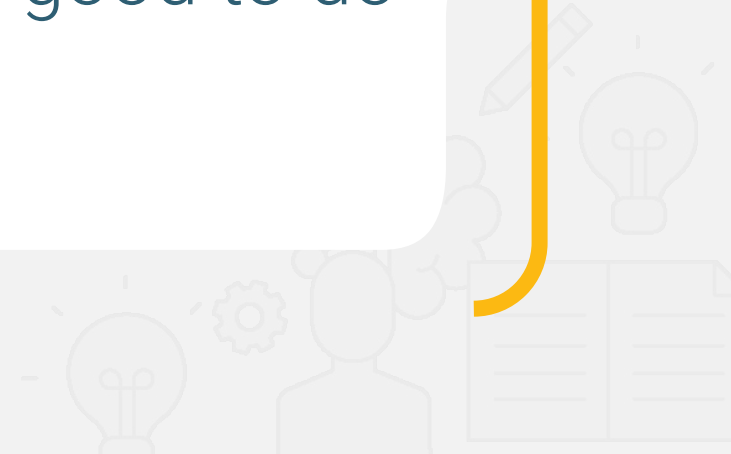
Welcome

Instructor **Demonstration**

AI Ethics

What is the first thing you think of when you hear the term **AI Ethics**?

Ethics are systems of principles that determine how people make decisions about what is good to do and bad to do.

Ethics in AI

The application of ethical principles to AI systems and technologies

Concerns about the impacts of these systems are cropping up at the same breakneck speed as AI technology development.

Although these issues are complex, an awareness of the major ethical concerns within the AI sphere can help guide the responsible creation and application of these technologies.

Significant Areas of Concern

01

Privacy

02

Consent

03

Accountability

Activity:

AI Ethics Exploration

In this activity, you will research an ethical AI issue of your choice.
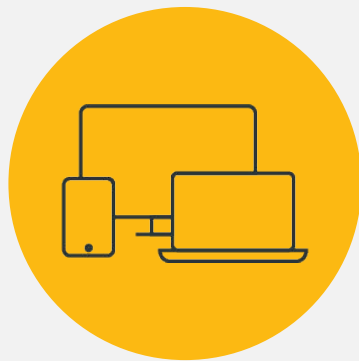
Suggested Time:

15 Minutes

Time's up!
Let's review

Questions?

Instructor **Demonstration**

Algorithmic Bias

Algorithmic Bias and AI

Algorithmic bias refers to **situations in which a computer system makes decisions that impact different people or groups of people in different, unequal ways**.
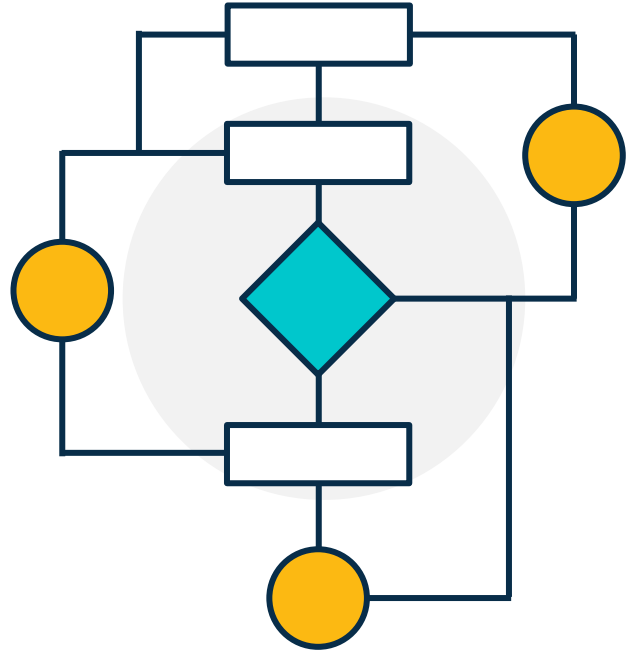
AI models are particularly susceptible to algorithmic bias because they're trained on massive amounts of existing data.

Algorithms

Algorithms are sets of logical, sequentially ordered steps.

Whether the algorithms are simple or complex, they are frequently applied to sets of data to process that data in some way.

Types of Algorithms

Prioritization

Ranks data based on factors that the algorithm design specifies

Classification

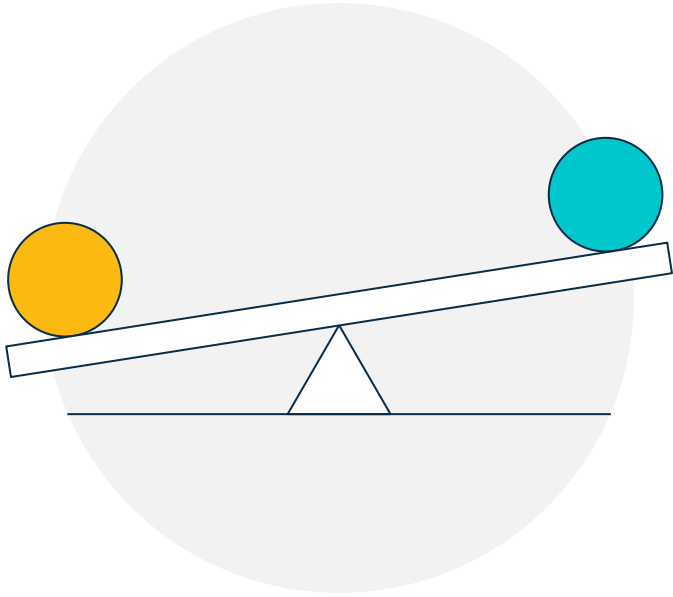Sorts information into categories

Association

Connects pieces of information typically related or used at the same time

Filtering

Removes or separates information that meets a certain criterion

Bias

Bias is usually defined as a situation in which one group, person, or thing is treated differently than other groups, people, or things.

People often discuss bias in terms of fairness or unfairness. Additionally, this discussion often occurs in the context of highly personal experiences of bias related to gender, race, or class. But bias can exist in many contexts.

Think Through All the Decisions You Made Today

How did you choose to get to work?

Did you shower, take a bath, or did you just wash your face when you were getting ready this morning?

What did you have for breakfast?

- **How long did it take you to make these decisions?**

- **Can you identify the factors that went into them?**

Identifying Biased Decision-making Can be Challenging

It's not always easy to know why we make the decisions we do.

We often have to make decisions quickly, and we might have to choose between options that seem basically equal.

In these situations, we tend to make quick decisions based on our past experiences. And, we're not always aware of the rationale behind our decisions.

Unconscious bias occurs when we make biased decisions based on factors we're not aware of.

Activity:
Automated Disqualification from SNAP

In this activity, you will read a series of articles about three Somalian markets in Seattle that were disqualified from the Supplemental Nutrition Assistance Program (SNAP) in 2002, forcing customers to shop elsewhere.

Suggested Time:
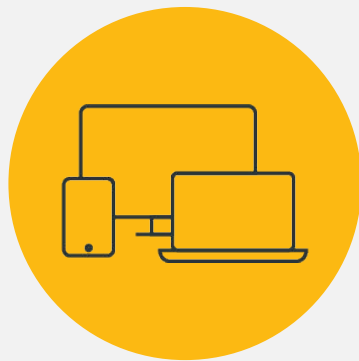15 Minutes

Time's up!
Let's review

Questions?

Break

15 mins

Instructor Demonstration

Causes of Algorithmic Bias

Causes of Algorithmic Bias

It can be difficult to determine the causes of algorithmic bias.

Two root causes are:

1 Algorithms are developed by large teams of company employees, making it difficult for anyone to take personal responsibility for ensuring that they are unbiased.

2 Humans have limited insight into why AI algorithms make the predictions they do (aka the "black box" issue).

Two Primary Factors that can Cause Bias

01

The people who developed the algorithm come from different backgrounds than those impacted by the algorithm.

02

The algorithm was trained with biased training data.

Recall: Exclusion from SNAP Benefits Activity

1   Minority groups are less likely to be represented on development teams.

2   If minority groups behave differently than the majority or different to what developers anticipated, their behaviors might be highlighted as abnormal or suspicious or be unaccounted for by the algorithm.

3   As unconscious bias can easily cause developers of an algorithm to make decisions that favor their own backgrounds and don't account for others', this can lead to algorithms that perform better for people like those developers and less well for others.

Biased Training Data

Training data refers to a labeled dataset that's used to teach an algorithm how to make decisions.

The algorithm bases its decisions on the patterns that it identifies regarding how the data points match their labels.

If the data is skewed in some way, i.e., it over represents a particular group and under represents another, the trained algorithm will perform better when a given data point resembling something from the first group and worse with the second.

If the data comes from an already biased source, i.e., if an automated hiring algorithm is trained on hiring data from a company that has historically prioritized men over women, the trained algorithm will learn to reproduce that bias in its decision-making.

Activity:
Healthcare Bias

In this activity, you will learn about an instance of algorithmic bias in the healthcare industry.
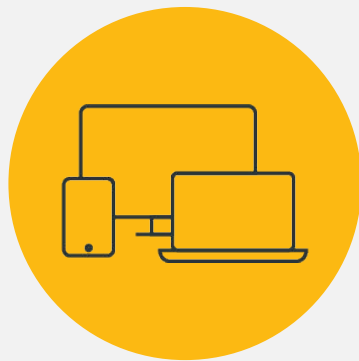
Suggested Time:
15 Minutes

Time's up!
Let's review

Questions?

Instructor **Demonstration**

Addressing Algorithmic Bias

Three Possible Strategies for Avoiding Algorithmic Bias

01

Auditing

- Proactively review algorithms to seek out bias.
- This can be done internally or externally.

02

Transparency and Documentation

- Make algorithms more transparent so that it is easier to point out overlooked areas of bias.
- Developers should be open about:
  - The source of data being processed
  - The specific information the algorithms uses and how it prioritizes information
  - Standards for what to expect as a result from the algorithm

03

Contestable Outcomes

- Build in mechanisms to provide feedback and contest results.
- For example, use thumbs-up or -down systems for recommendation algorithms

Checklists for Avoiding Algorithmic Bias

Existing systems

❏ Investigate how the system works. Was it trained using historical data? If so, think about biases that the historical data might contain. Also, seek out evidence that the developers have controlled for biased training data.

❏ Research similar systems to find out if they include examples of bias.

❏ Try auditing the system by testing it to find out if the accuracy of its results vary according to your input.

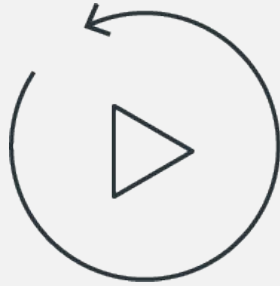❏ Check that a process for contesting the results of the system exists.

Systems in development

❏ Consider the types of people your system will serve or impact. How do these people compare to those who are developing the system? Make sure your team is aware of any differences, and teach them to seek out potentially biased decisions.

❏ Carefully review your training and testing datasets to find out if they under-represent or over-represent any groups of people.

❏ Consistently document your work in plain language so that others in your organization can both review and understand it. As much as possible, choose modeling techniques that people can understand.

❏ Create an internal process for auditing your system. And, identify someone outside your development team who can test that process.

Questions?

Let's **recap**

Recap

In this lesson you learned how to:

1 Explain what ethics means.

2 Summarize the major ethical concerns related to AI.

3 Explain why algorithmic bias is such a significant ethical concern.

4 List some common causes of algorithmic bias.

5 Analyze a case study in which researchers identified algorithmic bias in a system.

Next

In the next lesson, you will learn about the ethical and legal issues of working with AI, with an emphasis on data collection.

Questions?

The End