Module 1: What is Human Centered Data Science Slide Notes

Slide 1: Title

Title: What is Human Centered Data Science?

Slide 2

Title: Class Objectives:

* Define human centered data science
* Explain what does “human centered” add to data science
* Introduce methods of Human Centered Data Science and the data science cycle

Slide 3

Title: Myths and Assumptions about Data Science:

Definition: We can make a lot of assumptions about data science, about its potential to improve peoples lives, to prevent harms of the past, and to shape a more automated and fair future.

-These assumptions are founded on the idea that algorithms and data science are unbiased, that data is inherently neutral and unbiased

- It is often said that removing humans from decision making removes bias, or that using algorithms removes human error.

Example:

Slide 4

Title: Realities about Algorithms and Data Science

Explanation: Often people make the assumption that processes where they cannot see human input are unbiased or removed from inherent human biases. This is not true, in fact the data, and algorithms are encoded with bias from those who create them.

* Algorithms include societal bias and can exacerbate their impacts by encoding them into data structures and systems
* Data sets utilized by algorithms and as training sets are by their very nature biased, the hidden nature of this bias makes it insidious

Example:

Slide 5

Title: “Example: Algorithmic Screening in Child Welfare”

* Child welfare agencies use algorithms to flag which hotline calls should be investigated.
* The model can encode and amplify existing racial and socioeconomic inequalities.
* Human workers, institutional policy, and data all shape how ‘risk’ is defined and acted on.

Example of Algorithmic Bias

Slide 6

Title: How to address algorithmic bias:

Definition: Algorithmic bias is when

* To understand and prevent algorithmic bias data scientists need to understand the purpose of their algorithms,
* who may be harmed by them and recognize downstream impacts of using these algorithms.
* It is important to make transparent how algorithms are built and how the decisions they are making are being determined.

Algorithmic bias occurs when...

To effectively address and prevent algorithmic bias, data scientists must:

* Clearly define the purpose of their algorithms.
* Identify potential harm to individuals or groups.
* Recognize the long-term, downstream consequences of the algorithm's deployment.

Furthermore, ensuring transparency is vital: it must be clear how algorithms are constructed and how they arrive at their decisions.

-Example:

Slide 7

Title: How to address data bias:

-When addressing bias we need to begin by recognizing who is included in the data and who isn’t, think about how the data was collected, and how the data will be used.

-Each of these considerations help us to understand the potential for encoding bias, and to recognize who may be harmed by the decisions we are making.

-Example

Slide 8

Activity:

Turn to someone next to you and discuss a time when a social media algorithm presented you with content that seemed unrelated to your usual user behaviour and what may have contributed to this recommendation.

Slide 9

Title: How do we Define Human Centered Data Science

Human-centered data science is:

* a multidisciplinary field of study that seeks to develop awareness of the complex nature of the interaction between society, technology, and human-generated data.
* That rejects the notion that automated or computational approaches can be “free” of bias and shows how the power to define categories and ask questions has huge implications for the work that we do.

Slide 10

Title: What is the Purpose of Human Centered Data Science

-The purpose of human-centered data science methods is to address concerns about the

Social impacts of large scale data.

-Human Centered Data Science adds the human element and the consideration of human impacts back into data science. Shifting the lens from what is possible to what is possible, who does it help and harm, how can I ensure that what I am doing is a net benefit to all those impacted.

Slide 11

Title: Steps to Centering Humans in Data Science

* It’s important to understand human and social contexts; this helps data scientists respond ethically to the problems they are addressing.
* Human Centered Data Science goes beyond just usability and asks about impact, working to develop empathy for those impacted by data science.

Slide 12

Title: Why We Must Recenter Humans in Data Science

-A growing concern around data science is that without seeking a human centered approach data science may be so harmful that it does not meaningfully succeed in its goals of creating a better future.

Question: Do you think that data science can make a better future for everyone? Why or why not? Have your thoughts on this changed? Why or why not?

Slide 13

Title: The objectives of human-centered data science:

* human-centered data science endeavors to build rigorous and ethical algorithms that focus on the people who are involved in data science at every level.
* Question: What does it mean to meaningfully reflect on the responsibility that comes with the possibilities of data science? Is the act of reflection alone enough to guard against potential harms?

Slide 14

Activity:

As a small group create a flow diagram of the people across the data life cycle, may be helpful to use a specific data scenario for this activity (ex. Mortgage application algorithms, Ai for job applications, Ai for rental apps)

* think about the data itself
* Who is transforming the data?
* Who is impacted after algorithms are created?
* What are the potential long-term benefits or harms

Slide 15

Title: Methods of human-centered data science:

Definition: human-centered data science adds to the standard methods of data science by recentering the humans involved at every stage in the data science life cycle.

* Conducting qualitative research, including data collection and data analysis, that focuses on studying the people who produce data sets, algorithms, run data analysis, and interpret results.
* Example/question

Slide 16

Title:Human-centered data science as ethical responsibility

Definition: Decisions made at every step of the data science cycle require doing the work responsibly and treating everyone affected with care.

* Each step of data science relies on assumptions at best and bias at worst
* Methods to detect and mitigate bias will evolve overtime
* Large scale data presents challenges to rights and privacy, even innocuous data can quickly become identifiable

Slide 17

Title: Human-centered data science looking in the right places

Definition:

Framing the right questions and looking for their answers in the right places

* Focus on a goal for the project, not just look for a convenient answer
* Select the best data to evaluate these problems not just the simplest to get
* Social media with accessible API’s can make it easy to aggregate “perception” but may not be the best place to look for this information

Slide 18

Title: Human-centered data science as collective practice

Definition:

People from many different backgrounds can be effective data scientists given the right tools

* Working with domain experts is key to doing human centered data science
* Data science is collaborative and should include stakeholders
* When working on a project are you working with a team? Who all is a part of this team and what perspective do they bring?

Slide 19

Title: Human-centered data science as communication

Definition:

Human centered data science focuses on real world impact, how we communicate these findings is central to ethical and responsible data science.

* Data science can be story telling, how do we communicate to those outside of our team
* Good data science is collaborative and relies on good communication across a data team
* Micro Example/question

Slide 20

Title:Human-centered data science as action

Definition: Data science systems can privilege one group over another, we must be clear in what benefits and what harms these systems create.

* These systems can create winners and losers
* Data science systems can have life altering impacts
* With whom do you believe your lot is cast?

Slide 21

(with icon)Title : Exit Ticket

Activity:

Make a slide showing what you think is the data science life cycle

Individual Exit Ticket:

Multiple choice

1. What is not an assumption of human centered data science?
   1. **Algorithms are not by their nature biased**
   2. Data sets contain bias because of who is represented and who is not
   3. It is important to select the right data set for a data question not the easiest to find
   4. Data science questions should be crafted around the why
2. Which of the following is a human centered data science method?
   1. Data visualization
   2. Machine learning
   3. **Contextual understanding**
   4. Data collection and labelling
3. Which of the following are people in the data pipeline?
   1. End users/subjects
   2. Populations represented in the data
   3. Data scientists
   4. **All of the above**
4. Which is not an example of possible bias in data science?
   1. Historical data missing certain populations
   2. Algorithms built to mimic current processes
   3. Expectations of data scientists of the correct outcomes
   4. **Statistical error**

5.Reading