

---

CS 232:  
Artificial Intelligence

Spring 2024

---

Prof. Carolyn Anderson  
Wellesley College

# Reminders

---

- ❖ My Friday help hours are canceled, but feel free to email me with questions
- ❖ HW 9 is extended to Thursday, but try to submit by then so that I can give you feedback on your dataset in time for you to tweak before your final presentations.

# Bonus Late Day Opportunity

---

An Afternoon with Annabel Rothschild '20



3:45-4:45pm

April 22nd

L045

Annabel Rothschild  
Georgia Tech

# Bonus Late Day Opportunity

---

## AI for Wireless and Wireless for AI: A Tale of Two AIs



4-5pm  
April 23rd

Francesco Restuccia  
Northeastern University

Recap

# Stakeholders

---

There are different kinds of stakeholders to consider when we talk about the ethics of AI (Bender 2019):

- ◆ **Voluntary direct stakeholders:** people who choose to use the system.
- ◆ **Involuntary direct stakeholders:** people who must use the system in order to access essential services.
- ◆ **Indirect stakeholders:** subjects of queries, contributors to a corpus (voluntarily or involuntarily)
- ◆ **Project funders:** the people providing the funding
- ◆ **System builders:** the technologists creating the system
- ◆ **Communities:** communities impacted by model predictions

# Kinds of Harm

---

- ◆ *Allocational harms: Does the system allocate opportunities or resources unfairly? Do some people gain access more easily than others?*
- ◆ *Representational harms: Does this strengthen stereotypes? Does this create or reinforce unfair negative perceptions of a group of people? Does the system fail to even recognize some people?*

# Community Jury Roleplay



# Ethics assessment model: community jury

---

In the **community jury** model, the potential harms and benefits of a proposed technology are weighed by a diverse group of stakeholders.

- ◆ The **product team** creates relevant documentation, data management plan, and prototypes to present.
- ◆ A **moderator** facilitates discussion and deliberations.
- ◆ A **jury** is assembled of 16-20 community members, sampled in a way that is random but ensures a demographically diverse group.

# Ethics assessment model: community jury

---

2-3 hr sessions are held to assess the proposed technology:

- ◆ **Overview and introduction:** The moderator explains the rules of conduct. The product team explains the proposed technology and its goals.
- ◆ **Q&A:** jurors ask questions about the technology.
- ◆ **Deliberation and cocreation:** the jury and product work together to come up with solutions that meet all needs.
- ◆ **Anonymous surveying (optional):** anonymously poll the jurors to get their honest opinions.
- ◆ **Study report:** the moderator writes a report outlining key insights, concerns, and proposed solutions.

# Scenario: Code Generation

---

Roblox, a platform where people can program and share games, is collecting code to train a large language model of code. Their goal is to improve the experience of novice programmers.