AI Boot Camp

Legal and Ethical Issues in AI

Module 15 Day 2

Class Objectives

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

1. Describe the concept of data privacy and how it relates to AI.
2. Explain the concept of consent and why it is important for AI projects.
3. Outline the key legal parameters governing AI training data.
4. Summarize some legal considerations for selecting AI training data.

Instructor **Demonstration**

Data Privacy

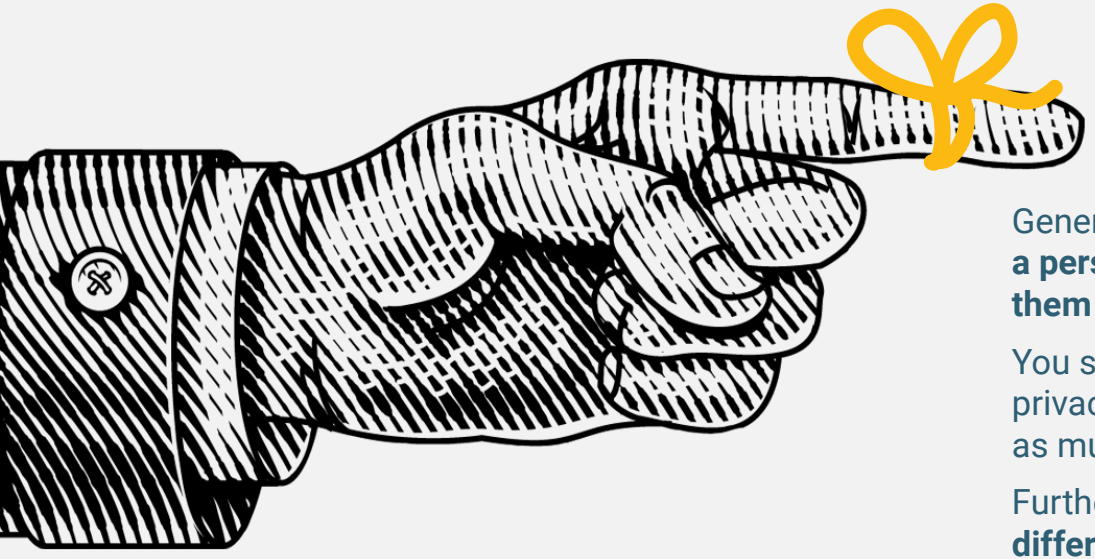
Think About Privacy in Your Life

1. In which circumstances in your life do you want privacy?

2. When was a time that you felt your privacy was violated in person? How about online?

3. When was a time that you might have violated someone else's privacy?

Keeping your answers to the previous questions in mind, how would you explain the **concept of privacy** to someone else?



Remember,

Generally speaking, you can think of privacy as **a person's right to control how information about them is shared.**

You should always be aware that the bounds of privacy will differ in **different contexts**, and that it is as much of a social issue as a legal one.

Furthermore, your **expectations of privacy** may be **different** than your friends', as well as anyone who is from a different geographic, socioeconomic, or cultural background.

Personally Identifiable Information (PII)

PII refers to information that can pinpoint the identity of a specific person, including:

- Full names
- Personal identification numbers, such as Social Security number, driver's license number, or passport number
- Directory information, such as address, email address, and telephone number
- Technical identifiers such as IP address
- Personal characteristics, such as an image of a person, recording of a voice, or other biometric information
- Information that can be linked to any of the above, such as date of birth, race, employment information, and education information

Activity:

How Unique are You?

In this activity, you will use the “How unique am I?” tool created by Dr. Sweeney’s Data Privacy Lab at Harvard.

Suggested Time:
10 Minutes

Time’s up!

Let’s review

Questions?

Instructor **Demonstration**

Consenting to Share Your Data

Consent can be understood as **agreeing** for something to happen.

Consent and Control

Consent in our context, gives people **control** over what happens to their information.

We exert this control in one of two ways:

- **Providing consent** for our information to be used in a certain way

- **Refusing consent** for our information to be used in a different way

Informed consent requires that subjects not only provide permission, but also **understand precisely what that permission entails**.

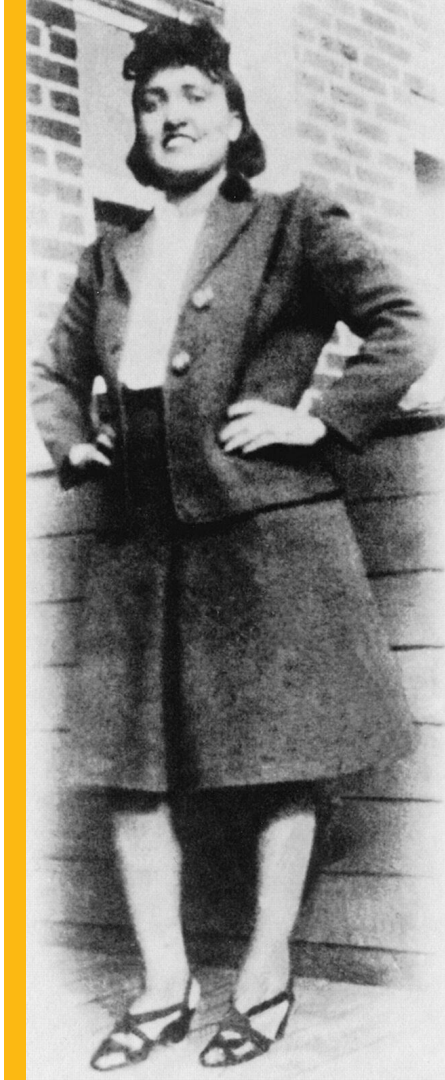
Consent Issue 1: Genetic Information

Henrietta Lacks' cancer cells were taken in 1951 without consent. They've been used for medical research for decades, continuing to this day.

- Her cells have been replicated indefinitely.
- Her genome has been sequenced.
- All of her future relatives can be easily identified through the harvested genetic information.

While this has lead to numerous medical advancements, it has also enriched pharmaceutical companies and brought fame and fortune to scientists and research organizations without an acknowledgement of Henrietta's part in the process.

Because Henrietta Lacks' family was never given an opportunity to consent to the removal and use of her cells, they were not able to share in the profits or to celebrate the impact their genetic line had on the world.



Consent Issue 2: Facial Recognition

In November 2021 Clearview AI was fined 22.6 million dollars for using images collected via web scraping to train a facial recognition algorithm without the pictured individuals' consent.

These participants have no way of knowing how this technology will evolve or what impact their inclusion in the dataset might have on their lives.

- What if being included in this dataset reveals who someone is and that results in sensitive information being leaked?
- What if someone didn't consent to being included in this dataset?
- What if they did consent, but they didn't understand how that information might ultimately be used?

Recommendation Algorithms

Concerns regarding consent for the use of personal data within AI stems from the opacity or “black box” nature of the technology.

This means that it is often impossible to know how algorithms are using the data they are fed to make decisions.

Recalling Cambridge Analytica

Consulting group Cambridge Analytica scraped the personal Facebook data of millions of users to train models. The model was used to politically target users and push them towards voting for particular candidates, in this case Donald Trump, in elections.

This scandal revealed not only how easily private user data could be accessed by third parties, but also how well it could be wielded in manipulative ways with little to no awareness by those affected.

Activity:

Clearview AI

In this activity, you will review and discuss the ways that Clearview AI obtains and uses information about people for its facial recognition tool.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Questions?

Break

15 mins

Instructor **Demonstration**

Who Isn’t in this Data?

Exclusion and Discrimination

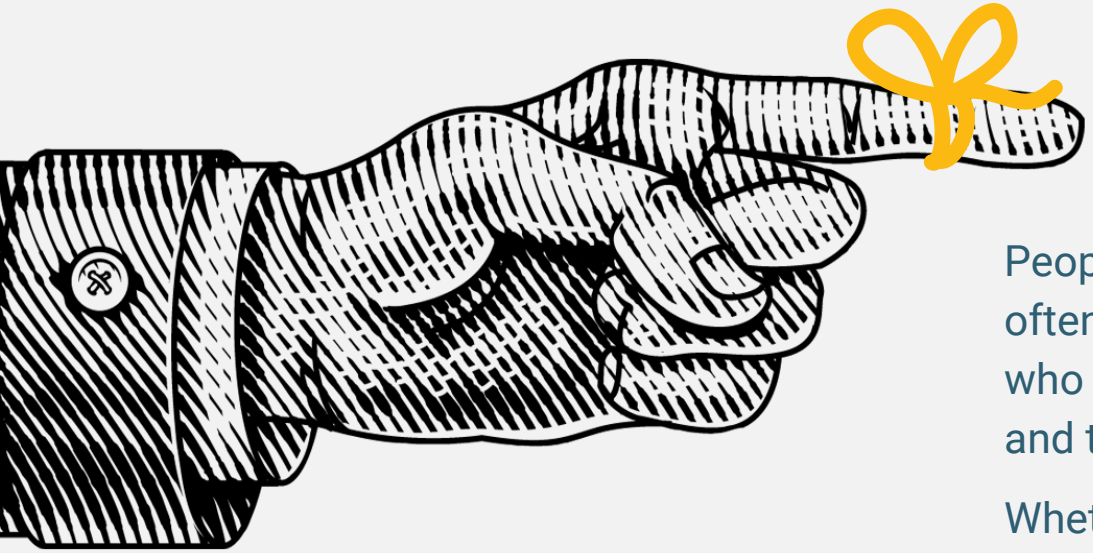
When working with datasets for the training of AI algorithms, it is important to begin by asking the following questions:

01

Who will be impacted by the decisions made based on this data?

02

Are all of those people represented in this dataset?

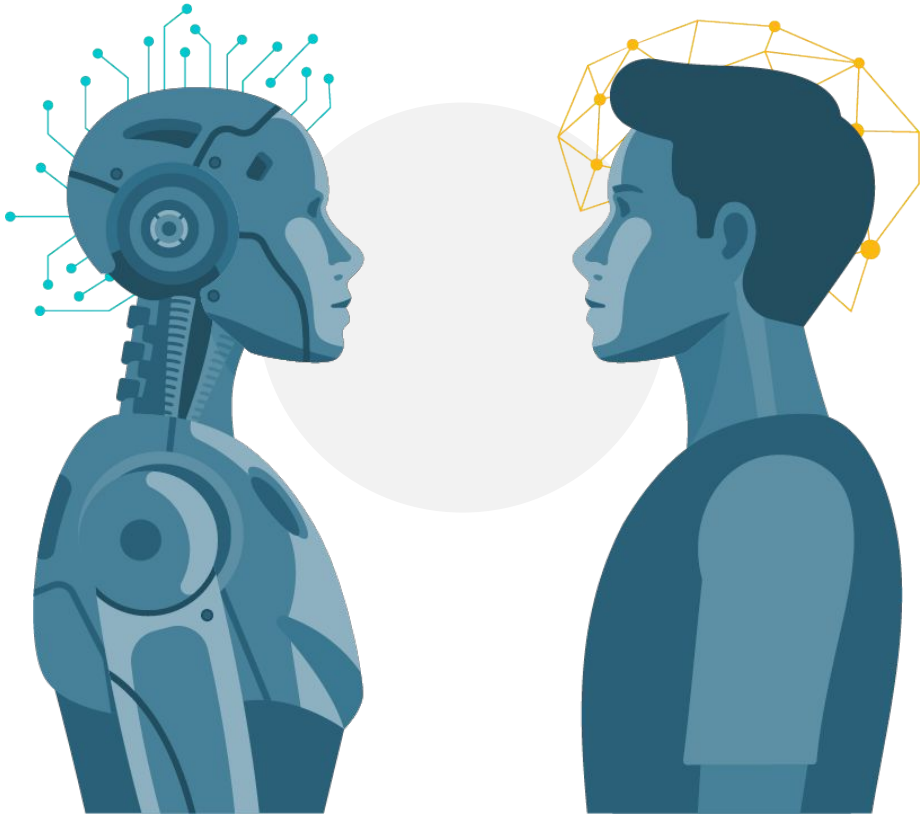


Remember,

People working on AI and AI ethics are often not representative of the groups who will be impacted by the technology and their decisions.

Whether intentional or unintentional, the results of exclusionary practices are the same.

Exclusion and Discrimination (cont.)



The perceived improvements that computer-based decision-making has over human decision-making complicates our ability to notice and acknowledge built-in biases.

Because it can be hard to determine precisely how an algorithm came to a particular decision, excluding data labels that carry the risk of discrimination does not prevent an algorithm from using related shared traits to reproduce the same biases.

Checklist for Ethically Using Data

- Review your dataset for any personally identifiable information (PII).
- Investigate how your dataset was collected.
- Think about who the people represented in your dataset are, and how they relate to the people who your work is likely to impact.

Activity:

Project Street Bump

In this activity, you will learn about an instance of exclusion bias from an example of an initiative designed to help fix potholes in the city of Boston.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Questions?

Instructor **Demonstration**

Copyrights and Contracts

What is Copyright?

Copyright grants authors of “original work” the automatic right to determine how their works can and cannot be used.

“Original works” include but are not limited to:

1. Literary works, such as books
2. Musical compositions
3. Dramatic works, like plays
4. Choreography
5. Visual art, including photography
6. Audio and video recordings
7. Architectural designs

However these rights do not extend to the “ideas, concepts, or principles” used in or introduced by the works in question.

Can Databases or Datasets be Under Copyright?

This is the subject for legal precedent.

Precedent:

Decisions made in previously litigated cases that are used as an authority for interpreting the law in future cases.

Consider Feist v Rural:

The court case ruled that Rural's subscriber list was a collection of facts and was therefore not copyrightable.

Contracts and Licenses

Contracts are agreements that can be enforced by law. These include contracts for large purchases and employment arrangements, but are also included in the use of software, websites, and datasets.

End-user license agreements, **terms of use**, or **terms of service** are contracts that most companies require you to agree to in order to use their products.

In most cases these have been ruled to be legally enforceable protections of particular databases and datasets. But there are notable exceptions such as in the case of ProCD v Zeidenberg.

Web-scraping is the process of extracting information from websites. It can be a powerful way to gather information for the assembly of large datasets.

In many cases, datasets created by using web-scraping are legal to use, but it remains a legal gray area as regulatory bodies struggle to catch up to advancing technologies used to automate the process.

Are you Web-scraping Legally?

1. Are you accessing a computer?
2. Are you authorized to access the computer in the way that you are accessing it?
3. Are you obtaining information from a protected computer?

Can generative AI be **ethically trained** on existing works?

Activity:

The Legality of Generative AI

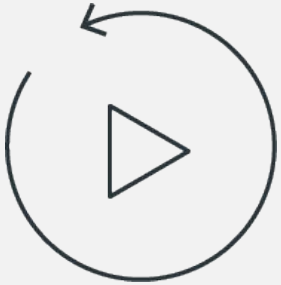
In this activity, you will research a legal case of your choice that revolves around generative AI.

Suggested Time:
15 Minutes

Time’s up!

Let’s review

Questions?



Let’s **recap**

Review the Class Objective

In this lesson you learned how to:

1. Describe the concept of data privacy and how it relates to AI.
2. Explain the concept of consent and why it is important for AI projects.
3. Outline the key legal parameters governing AI training data.
4. Summarize some of the legal considerations for selecting AI training data.

Next

In the next lesson, you will learn about the laws and regulations that govern AI.

Questions?

The End