Data Science Ethics

College of Health and Medical Techniques

Al-Mustaqbal University

Stage 2 , Semester 1
@ Department of Intelligent Medical Systems

Prepared by
Dr. Samir Badrawi

Course Preview

The majority of this course material is based on Coursera
https://www.coursera.org/learn/data-science-ethics
"H.V. Jagadish lectures", a Professor at the University of Michigan

What are Ethics?

The word "Ethics" is derived from the Greek word Ethos, meaning habit.

Ethics tells us about right and wrong. Our course is about the ethical practice of Data Science.

Ethics are shared values that help differentiate right from wrong.

Ethics are not law, but they are usually the basis for laws.

(we will more elaborate the definition of "Ethics" later)

But what is Data Science?

Data Science is an interdisciplinary field that involves the extraction of knowledge and insights from large, complex datasets. This is done through a combination of various techniques, algorithms, processes, and systems.

Data Science combines expertise from several domains: statistics, computer science, machine learning, …. to analyze and interpret data in order to solve complex problems and make <u>informed decisions</u>.

1. **Ethics and Data Science - Course Preview**

Today, we live in the age of **big data**. We live in a world that people couldn't have imagined a generation ago.

Your work will be concerned with data, and as "**data scientist**" you should think about the ethics of working in the field of Data Science.

Cont.

If I take a **picture of you**, who owns that picture? In terms of law and in terms of our general understanding, that picture is a picture of you, but I took that picture. I own it.

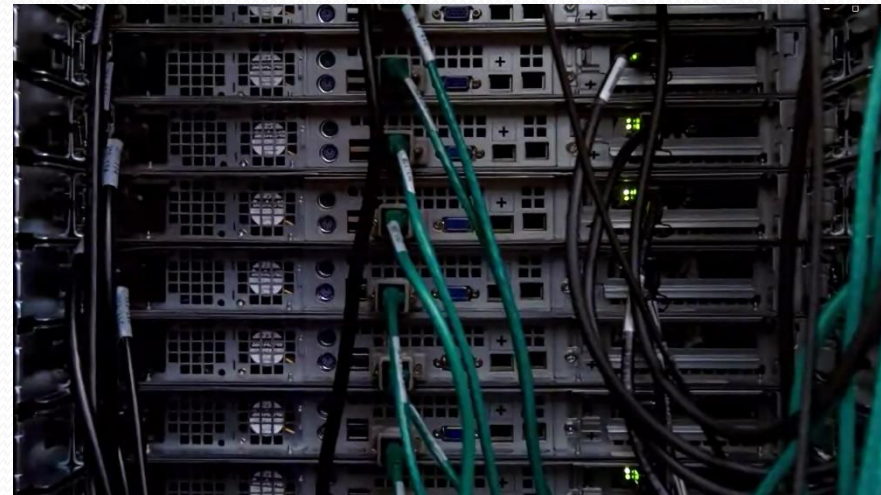**What are ways in rechecking to use that picture, and whether I can use that picture in ways that could hurt you?**

Cont.

Now take this to data:

If I have some data about you, is that your data or is that my data because I collected it?

Cont.

Just trying to understand what's my ownership of this versus what's your ownership, and ..
what's the obligation one has to the other is not so straightforward.

What one needs to have first is a societal agreement on what is right and what's wrong.

And then, flowing from that, a framework to think about how they could practice their data science differently so that all do the right thing.

Cont.

If you're doing things with data, you got to remember that you have a great deal of power.

And with a great deal of power, comes a great deal of responsibility.

Cont.

Data scientists who have had ethical education and training will be in better more ethical practice of data science, and that this is good for society at large.

In this course, we will learn how be responsible in that exercise of power.

2.What are Ethics?

Basic definition of Ethics

Rules that we all voluntarily follow because it makes the world a better place for all of us.

They help us distinguish between right and wrong.

They are the cornerstone of civilization.

If we don't have ethics, we're not civilized.

- **Ethical principles stopped me from stealing your wallet.   (No Theft)**

If you're bigger and stronger, I may not want to steal your wallet !

But ethical principles stop me even
     (1) if you're smaller and weaker
     (2) if there is no chance I will be caught

- **<u>Ethics are not Religion</u>**

Most religions promote ethical behavior.

(Thou Shalt Not Steal)

But ethics need not to be religious.

Ethics flow from shared values, which could be on account of religion, or not.

- **<u>Ethics are not Law</u>**

Suppose you tell me a secret, and I promise to tell no one. I then break my promise.

I may **<u>not have broken any law</u>**. It's unlikely that I will be punished for having broken my word to you. But I think we would all agree that in doing so, I would have been **<u>unethical</u>**.

Cont.

Well, ethics are not laws

But ethics guide the creation of laws, so the two are often in consonance.

The problem is that, even if we have shared values, not everyone will be ethical.
So, Laws may be used to enforce ethical behavior ... that is why people go to jail foe theft.

- **<u>Economic benefits of Ethics</u>**

for the most part, society as a whole does best when each individual works to maximize their own individual benefit.
This is the basis of the free market that underlies so much of modern economy.

But there also are situations where individual benefit comes at a cost to society.
And when this happens, **<u>shared societal value systems</u>** are required to address these situations.

Cont.

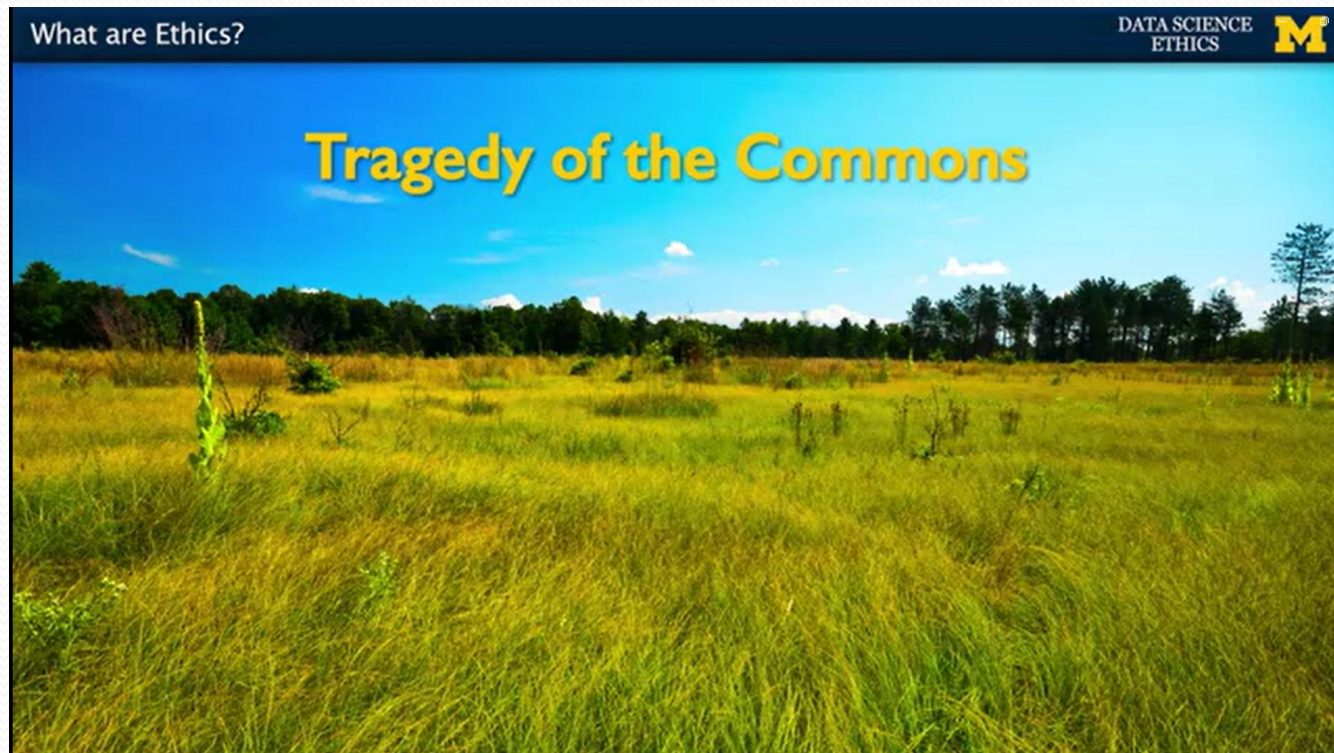## individual benefit versus (vs) society benefit

At buffet restaurants, customers pay one flat price and then help themselves to an array of different foods arranged on hot and cold food tables.

Cont.

In terms of economics textbooks, there is something called the tragedy of the commons

There is an open field that is the commons.
It's shared across the entire village.

Cont.

Now, since this is a **shared field**, it's fine for me to have my sheep graze in the commons.

That way I have less pasture that I have to buy, and I can afford to keep more sheep because my sheep can make use the grass in the commons as well as on my own land.

Cont.

But if everybody in the village makes that same judgment, then the <u>land in the commons gets overgrazed.</u>

Nobody takes care of it and we end up with a wasteland instead of having a lush meadow that we would all have enjoyed.

Cont.

Cont.

The only way to avoid this is for us as a village to come together and <u>agree on some rules</u> with respect to how we can all benefit from the commons without over exploiting it.

So, to summarize, a rule is something that limits what I can do, and therefore there's a cost associated with me following the rule. But I might still like to have a rule in place if the cost of me of following this rule is less than the benefit to me if others follow the same rule.

This is the thing that causes us to have an economic benefit from these shared values that drive rules that regulate our behavior.

- **<u>Let's consider a simple example</u>**

Suppose we had **<u>no rules</u>** about how we would drive well, we'll have **<u>chaos</u>** on our roads.

But we all see the benefit of all of us agreeing to a particular rule and agreeing to drive on a particular side of the road, and most of the time this rule results in us being able to get to where we want much faster than if we didn't have this rule.

- **Let's look at another example**

If you have litter and other people have littere, then you're not motivated to do a good job with your litter ?

- **<u>So to sum up</u>**

 ethics are shared rules that we all agree to follow because of the resulting benefits. These rules are typically infused with a sense of right and wrong.

In this first lecture we've learned why we're all better off with ethical behavior.

<u>In the next lecture</u>, we will link this up with **data science**.

3. Why Data Science need Ethics

- **The importance of ethics, for data science.**

  There's tremendous excitement about data science precisely because it gives us a better way of doing things in so many aspects of society.

  Data Science has impact because it has social consequences, **but this impact can also come with undesired consequences, for privacy, for fairness, and equity, so on**.

Cont.

As an example, **unsolicited email** was something that was considered a great idea in the 1990s.

Over time **spam** became a big problem.

No upstanding business will own up to spamming intentionally.

Cont.

This is just something that is considered socially unacceptable. (no good)

But, in a similar vein, much of the work that data scientists do has potential for delivering great values. It can help in some many different ways. (good)

So, Developing a **shared sense of ethical values** is necessary if you want to reap the benefits of data science, while minimizing the harm that it could possibly do.

- In 1994, the Internet was young, and there were a lot of clever people trying to figure out what to do with this new medium, and exploring the possibilities that it opened up.

- In particular, there was a pair of attorneys in Arizona, Lawrence Cantor and Martha Siegel, who came up with the idea of reaching out to a vast audience through this new medium, the Internet.

Cont.

These attorneys were immigration attorneys and they wanted people who would be interested in their services for a <u>green card lottery</u>.

They sent a message about this lottery and asking people to hire them, to help them win the process, and they posted it <u>to every place that they knew how to posted it.</u>

This is the equivalent of sending **unsolicited email** to everybody. This is the first known instance of what we now know as **spam**.

Cont.

The word **spam** was derived from a **Monty Python skit** TV show.

Spam was a *brand name for canned, processed*

A clip in one of the episodes featured Vikings who passionately chant about Spam, the brand name for canned, processed meat. Their chants in the clip became so loud and repetitive that they muted any other conversations.

Let us see the following clip :

Eventually, this clip became synonymous with annoying messages, and "**spam mail**" was born.