EXperimental Learning

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

# Big Data and Social Analytics certificate course



**MODULE 6 UNIT 1**
**Video 2 Transcript**

# MIT BDA Module 6 Unit 1 Video 2 Transcript

**Speaker key**

CK: Cameron Kerry

HY: Hapyak

CK: This video will look at some of the privacy practices in judgments to consider when doing work involving data about people especially when that work may have an impact on people's individual behavior.

Public opinion surveys and other research on attitudes towards privacy show that people express a lot of concern about how data is collected, and how much data is collected about them, and how that data is used. On the other hand, people don't take a lot of steps to protect data that are available to them.

This contradiction gets referred to as the privacy paradox. The answer to the privacy paradox appears to be that people are willing to share information based on a certain level of trust in the recipient, and a lot of that trust depends on context. That's a major reason that we're willing to share very sensitive information with healthcare providers, for example.

Establishing and maintaining trust is an essential ingredient of privacy. Let's talk about ways to do that. Here I remind you about the FIPPs – the seven fair information practice principles I laid out in an earlier video.

HY: The seven Fair Information Practice Principles

CK: If you want to go back to those you can click on this link.

00:01:46

A basic rule that summarizes the core of these FIPPs is: no surprises. That is really what the principles of control, transparency, and respect for context boil down to. Data shouldn't be used in ways that people wouldn't expect when they shared the data. One way to ensure that there are no surprises is by employing privacy by design.

Addressing privacy shouldn't be an afterthought; it should be part of the design of the research or the system from the beginning. Part of that design is to address what will be done about protecting the identity of individuals involved. In another section of the course Yves-Alexandre de Montjoye covers ways that people can be identified from data sets and some of the methods for protecting that identity.

Thinking ahead about privacy should include considered decisions about what data you need. That's what focused collection is all about. Big data need not mean indiscriminate collection of data. And

here my emphasis is on the word indiscriminate. Big data means analysis on an unprecedented scope and scale but still warrants discriminating choices about what data to collect; not just collecting data because you can and because it might be useful some day.

Security is a vital component of trust, it's part of data management. In Massachusetts, where MIT is located, organizations are required to have a written information security plan, and to be prepared for security incidents. That is a fact of life in a connected world. In the end though, many of the hard questions can't be resolved by law or regulation.

HY: Why do you think it's not possible to create concrete laws to police data privacy in this technological era?

Thank you for your reflection. Please continue watching to learn more about the topic.

CK: Even in the EU, with the most prescriptive privacy and data protection law, they allow the processing of data based on the legitimate interests of the controller. These interests have to be balanced against the rights and interests of the data subject, the person the data is about, but there's not a lot of guidance on how that balancing takes place.

00:04:22

At today's pace of innovation the law lags too far behind technology to answer many of the hard questions. Take for example the increasing use of sensors, an automated collection of data. The classic approach of notice and consent, explicit verbal consent just doesn't work in that environment. So the questions become not just what you must do but what you should do. In this connection consider that most research involving personal data is a form of human experimentation.

Even the most basic AB test, however harmless, is an experiment involving human beings. So, in this light, ethical standards that have evolved involving human-subject research are a guide. The founding document in this area is what is known as the Belmont Report issued in 1979. It laid out three principles. First, respect for persons. Treat individuals as autonomous agents and respect their right to determine their own best interests.

Second, beneficence. Do no harm. Maximize benefits and minimize harms, and systematically assess both. Third, justice. Treat subjects fairly and equitably. A restatement of the Belmont Report in 2012 added a fourth principle, respect for law and public interest. Pay attention to what the law is and be accountable for compliance. At the end of the previous video on privacy I posited a hypothetical situation of an institution that wants to be involved in this course, and raised a series of questions that you might want to ask.

I suggest now there's another set of questions to ask regardless of whether you're in the European Union or the United States, or any place else. Is this information that we really need? Is the information likely to be important enough compared to its impact on the privacy of the individuals involved?

How would I want my own data used and handled? Privacy leaders talk about being good stewards of data. In the context of privacy, being good stewards means focusing on the interests of the people the data is about. That is the challenge that you face in analyzing data involving people.