

Video Segment 5: Addressing Biased Datasets

NOTE-TAKING GUIDE

“There are some statistical things that you can do with the data to try to correct for bias, but there’s only so far you can go with imperfect data.” -Dr. Hamsa Bastani

Often data that is collected for research does not provide an accurate representation of a population or situation; it is biased. Dr. Hamsa Bastani says, “We can try to bring in datasets from diverse perspectives and stakeholders to build a more holistic representative view that will be more fair and equitable to everyone.” Take note of how the Wharton professors on the panel have done this in their research:

Sending medicine to Sierra Leone:

Human trafficking work:

Identifying echo chambers in media consumption:

Embracing imperfect measures in policing reform data:

Wharton professor **Duncan Watts** seeks to expose bias in journalism by building a huge database of news for researchers and journalism watchdogs to analyze. A next step in the process: building an **interactive dashboard**. Draw from the video segment and the vocabulary to explain what this means for data accessibility. What's the value of digging into the data?

Reflection: Bias in data collection produces biased machine-learning models, which can lead to more problems than solutions in addressing social issues. Shift your lens away from technology for a moment. Bias refers to prejudice in favor of or against a person or group in a way that can be perceived as unfair. Bias shows up not only in our attitudes, but also in our actions and the ways that we treat other people. Have you acted in a biased way or been impacted in some way by bias? Describe the situation and how it made you feel. How does bias, whether it be in raw data or people's attitudes, hurt society and innovation?

THE LANGUAGE OF BUSINESS

(Linked terms are defined in the Wharton Global Youth Video Glossary)

[Data](#)

[Statistician](#)

[Census Data](#)

Academic Research: Research conducted by faculty at a university like Penn that focuses on creating new ideas and perspectives. Academic researchers read all the existing studies and papers examining a specific topic or theory and then see where they can add value with new experimentation and insight.

Artificial Intelligence or AI: Intelligence demonstrated by machines (computers), rather than humans.

Behavioral Data: Data generated by, or in response to, a customer's engagement with a business; data analysts can look for patterns in behavioral data.

Beyond Business: An ongoing conversation at the Wharton School of the University of Pennsylvania that explores the most complex and pressing issues impacting individuals and organizations across the world. Each year Wharton Global Youth creates [Explore Business mini-sites](#) around different Beyond Business themes.

Bias: Prejudice against or in favor of one thing, person or group compared with another, usually in a way that is unfair.

Biased Data Views: Views on issues that are driven by data errors that occur when certain elements of a dataset are overweighted or overrepresented, or underweighted and underrepresented.

Bias in Machine Learning: When an algorithm produces results that are systematically prejudiced due to poor assumptions in the machine learning process – perhaps, for instance, the data is imperfect or biased.

Data Analytics (or just Analytics): The systematic computational analysis of data or statistics. It is used for the discovery, interpretation, and communication of meaningful patterns in data.

Dataset: A collection of data.

Dean Knox: An assistant professor of operations, information and decisions at the Wharton School of the University of Pennsylvania and co-founder of [Research on Policing Reform and Accountability](#).

Deep Web: The part of the internet, often where drug sales or other illicit activity occurs, that can only be accessed with specialized software, not through a regular search engine. It's also called the Dark Web.

Duncan Watts: A professor of operations, information and decisions at the Wharton School of the University of Pennsylvania and director of the Computational Social Science Lab at Penn, where he leads the [Penn Media Accountability Project](#).

Echo Chamber: An environment in which a person encounters only beliefs or opinions that coincide with their own, so that their existing views are reinforced and alternative ideas are not considered.

Erika James: Dean of the Wharton School with research interests in crisis leadership, workplace diversity and management strategy.

Human Trafficking: The use of force, fraud or coercion to get some kind of labor or commercial sex act. Examples include forced labor and sex trafficking.

Interactive Dashboard: A business data management tool that allows users to interact with data by tracking, analyzing, monitoring, and displaying key business metrics. Users can dig into an organization's operational information and filter it in different ways so that data can be viewed from different perspectives or in more detail.

Penn Media Accountability Project (PennMAP): An interdisciplinary research project of the Computational Social Science Lab at the University of Pennsylvania dedicated to enhancing media transparency and accountability at the scale of the entire information ecosystem.

Research on Policing Reform and Accountability: A group of researchers from different academic institutions (including Dean Knox from Wharton) who develop cutting-edge statistical techniques to measure racial bias in policing, evaluate policing policy reforms, and improve the performance of policing organizations.

Satellite Data: Images and earth observation data of the earth's surface and its atmosphere gathered by satellites in Earth's orbit.

Sierra Leone: A country in West Africa.

Stakeholder: A person with an interest or stake in something, especially a business. In many companies, investors are stakeholders, as are employees, managers and customers.

Statistics: The practice or science of collecting, analyzing and interpreting numerical data in large quantities.

The Wharton School: The Wharton School, also referred to as just Wharton, is the business school of the University of Pennsylvania, a private Ivy League university based in Philadelphia, Pennsylvania. Wharton is a world leader in business education and a hub for academic research in all areas of business.

EXPLORE MORE IN OUR ONLINE BUSINESS JOURNAL

[Learning to Be Color Brave](#)