

## Video Segment 4: Uncovering the Data that Drives Sensitive Research

*“We care about getting inputs from a wide variety of stakeholders.” -Dr. Hamsa Bastani*

### NOTE-TAKING GUIDE

Both Dr. Hamsa Bastani and Dr. Dean Knox partner with nonprofits, law enforcement and other entities to get access to data for their research. Using the video segment as well as the vocabulary below, list several **collaborators** that help to improve Professor Bastani’s machine-learning models and enrich Professor Knox’s data discovery.

1.

2.

3.

4.

5.

6.

In his work with **Research on Policing Reform and Accountability**, Dr. Dean Knox has said that we are “massively underestimating” bias in policing because we don’t have enough data or statistical analysis. And yet, with policing it’s difficult to study or know what there’s no record of, plus existing data can be contaminated when administrative records of police stops or arrests are produced through a lens of racial bias. Drawing from the video segment, in what ways might police-civilian interactions *not* reflect the true picture of discrimination? How does this represent the concept of **selection bias** (Hint: use your vocabulary list for help)?

**Reflection:** Dr. Dean Knox says that students in his classroom show a huge appetite for the importance of getting the statistics right. Do you value data-driven social impact solutions? Would you pursue a career related to social awareness and change, supported by analytics and machine learning? Why is data a game-changer? What social issues would you like to tackle with the help of rich data and machine learning?

## THE LANGUAGE OF BUSINESS

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

[Data](#)

[Public Policy](#)

[Nonprofit](#)

[Social Impact](#)

[Policymaker](#)

**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.

**Accountability:** This refers to taking ownership for your decisions. When it comes to environmental issues, for example, businesses need to be accountable for their actions; they need to accept responsibility for their decisions and the impact those decisions have on the planet. The same holds true for other social issues, including policing.

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

**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.

**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.

**Dataset:** A collection of data.

**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.

**Data Scraping:** A technique in which a computer program extracts data from another program.

**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 known as the Dark Web.

**Department of Justice:** A federal executive department of the U.S. government that upholds laws, keeps the country safe and protects civil rights.

**Discovery:** The pre-trial procedure in a lawsuit in which each side must exchange evidence and legal information.

**Discrimination:** When someone treats another person unjustly based on the groups or classes to which they belong. For example, people may experience discrimination on the basis of their race, gender, age or religion.

**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](#).

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

**Freedom of Information Request:** Under the Freedom of Information Act, U.S. citizens have the right to request access to records from any federal agency.

**Global Emancipation Network:** In an effort to end human trafficking, this nonprofit delivers cutting-edge data and technology to stakeholders around the world.

**Hamza Bastani:** An assistant professor of operations, information and decisions at the Wharton School of the University of Pennsylvania, who has been analyzing dark web data related to human trafficking to provide new insights into online sex-trafficking recruiting practices and, more broadly, illicit supply chains on online platforms.

**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.

**Machine Learning:** Using and developing computer systems that learn and adapt by using algorithms and statistical models to analyze and understand patterns in data.

**Machine Learning Model:** A program that has been trained to recognize certain types of patterns and make decisions from data it has not yet seen.

**Police Body Camera:** A camera worn by police that record interactions between community members and law-enforcement officers.

**Policing Reform:** A movement to transform policing culture and practices, eliminate racial bias and profiling in policing, and end police violence against citizens.

**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.

**Selection Bias:** The bias (prejudice against or in favor of one thing, person or group compared with another, usually in a way that is unfair.) introduced by the selection of individuals, groups or data for analysis in such a way that proper randomization doesn't happen. The result is that the sample under study is likely not representative of the population that is being analyzed.

**Social Scientist:** An expert in human society and social relationships, including economics.

**Stakeholder:** A person with an interest or stake in something, especially a business. Investors are stakeholders, as are employees, managers and customers.

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

**Suffolk County Anti-Trafficking Initiative:** An initiative to track and address human trafficking on Long Island, New York.

**TellFinder Alliance:** A web platform that connects its users to the data of the deep web, where human trafficking and other crimes take place.

**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 FUTURE OF THE BUSINESS WORLD PODCAST

[Sustainable Farming: Developing a Monitoring System to Protect Crops from Disease](#)