



LGST 242x/642x

Big Data, Big Responsibilities: The Law and Ethics of Business Analytics

Spring 2018

Overview

Significant technologies always have unintended consequences, and their effects are never neutral. A world of ubiquitous data, subject to ever more sophisticated collection, aggregation, analysis, and use, creates massive opportunities for both financial gain and social good. It also creates dangers in areas such as privacy and discrimination, as well as simple hubris about the effectiveness of management by algorithm. This course introduces students to the legal, policy, and ethical dimensions of big data, predictive analytics, and related techniques, and examines responses—both private and governmental—to address these concerns.

Instructor

Associate Professor Kevin Werbach
Department of Legal Studies and Business Ethics
673 Huntsman Hall
(215) 898-1222 | werbach@wharton.upenn.edu (best way to reach me)

Office hours: M 11-12, Tu 12-1, or by appointment

Learning Objectives

Good data-driven decision-making means not just generating solutions, but understanding how to use them. As we will see, some of the most sophisticated firms in terms of data science expertise have already experienced ethical lapses, public relations disasters, regulatory sanctions, and even legal liability. My goal is to help you develop skills to use analytics in the most responsible way, while remaining focused on business objectives. After completion of the course, you should be able to:

1. Identify where algorithms depend on human judgments or assumptions.
2. Describe legal rules and regulatory obligations, in the U.S. and elsewhere, that may apply to business analytics.
3. Evaluate claims that applications of analytics raise ethical or public policy concerns.
4. Develop thoughtful responses to concerns about the uses of data science.
5. After graduation from Wharton, don't destroy the world, crash the economy, go to jail, or all of the above. (Money-back guarantee not available.)

The course is non-quantitative, although a basic familiarity with data science techniques is helpful. Many of the questions it covers do not have simple answers, either because concepts such as privacy cannot be reduced to binary choices, or because the scenarios involved are still unfolding. A major goal of the course is to help you develop your own well-grounded viewpoints.

Materials

All readings are contained in a coursepack available through Study.net. Virtually all are also available through the hyperlinks provided in this syllabus.

Course Requirements and Grading

Exam (40%)

There will be one closed-book exam, which will be administered during the final session. It will test your comprehension of the required readings and the concepts developed in class.

Interactive Activities (40%)

We will do a mock trial and group negotiation. You will take on roles and attempt to develop a consensus to address concerns about business analytics. You will receive credit as follows:

- Individual prep sheets (10%)
- Group negotiation proposal, or explanation of your failure to reach consensus (10%)
- Individual reflections (20%)

Participation (20%)

This is an interactive course. Most class sessions involve discussion or small-group activities. You will be assessed primarily on the quality of your contributions to each class sessions, not simply your attendance.

I realize there are understandable reasons students miss class. There is no penalty per se for missing up to four sessions. However, five or six unexcused absences (meaning arriving or leaving more than fifteen minutes from the start/end of the class) will each result in a one-step reduction in your final grade (e.g. A- to B+). Being absent from seven or more classes will result in failing the course.

You don't need to tell me in advance, or contact me afterwards, to explain an un-excused absence (i.e., not involving a medical issue, obligatory university/athletic trip, or family/personal emergency.) If you believe your absence should be excused, please email me an explanation and, where possible, documentation.

Policy on Devices

This class will observe the Wharton MBA Program's policy on student use of electronic devices in the classroom. Violations will result in deductions to the class participation aspect of the course grade. Specifically:

- Phones must be turned off and put away.
- The use of laptops and tablets is not allowed unless pre-approved for educational purposes, or specifically authorized in class by the instructor.

Instructor Bio

Kevin Werbach is an Associate Professor in the Legal Studies and Business Ethics department. His research focuses on Internet policy, big data, and blockchain technologies. He was Counsel for New Technology Policy at the Federal Communications Commission during the Clinton Administration, served on the Obama Administration's Presidential Transition Team, and created one of the first massively open online courses (MOOCs) on the Coursera platform.

Syllabus

Where there are questions listed under a reading, be prepared to address them in class.

1. THE PROMISE AND THE PERIL

How might data science change the relationships among firms, customers, employees, other firms, and governments? What are some of the problems that may arise?

The White House, Big Data: Seizing Opportunities, Preserving Values (2014), pp. 1-10

- *What are the key innovations of big data and business analytics?*

Leo Hickman, How Algorithms Rule the World, The Guardian, July 1, 2013

- *Algorithms are cool! Why should we worry if they rule the world?*

Steven Finlay, Predictive Analytics, Data Mining, and Big Data: Myths, Misconceptions, and Methods (2014), chapters 1-2 (Optional)

- *An overview of business analytics, for students without a background in the area.*

2. "IT'S JUST MATH"

Algorithms rely on human decisions about how data are collected, analyzed, and used. Assuming otherwise leads to problems when things go wrong.

Pedro Domingos, A Few Useful Things to Know about Machine Learning, Communications of the ACM, October 2012

- *What are the ways that machine learning requires human judgement or intervention?*

Neil Richards and Jonathan King, Three Paradoxes of Big Data, 66 Stanford Law Review Online 41 (2013)

- *What practical problems arise from the three paradoxes the authors identify?*

Zeynep Tufekci, The Real Bias Built In at Facebook, N.Y. Times, May 19, 2016

- *Why was Facebook criticized for its Trending Topics?*
- *Why does Tufekci say that algorithms are not neutral? What does that even mean?*
- *If algorithms are biased, does that undermine the value of analytics in business?*

3. LIMITS OF ANALYTICS

Where are business analytics techniques likely to fail, and why?

Gary Marcus & Ernest Davis, 8 (No, 9!) Problems with Big Data, N.Y. Times, April 6, 2014

- *What are some of the common themes in the authors' list of problems?*

David Lazer et al, The Parable of Google Flu: Traps in Big Data Analysis, Science, March 14, 2014

- *Why was Google Flu Trends so accurate initially, and not subsequently?*
- *Should the failure of Google Flu make us skeptical about the potential of business analytics?*

Michael Luca et al, Algorithms Need Managers, Too, Harvard Business Review, January-February 2016

- *What is the role for business managers in overseeing the use of analytics?*

4. FAIRNESS

How can algorithmic decision-making result in unfair results? And just what is fairness, anyway?

Solon Barocas & Andrew Selbst, Big Data's Disparate Impact, California Law Review (2015), pp. 677-93

- *What are the ways that big data can produce bias or discrimination?*

Julia Angwin et al, Machine Bias, ProPublica, May 23, 2016

- *Does the report demonstrate unfair or discriminatory outcomes from the use of the COMPAS system for sentencing?*
- *What might explain the racial variations the researchers found?*

Suresh Venkatsubramanian, Algorithmic Fairness: From Social Good to a Mathematical Framework, LSE Media Policy Project Blog, June 14, 2016

- *Can technical solutions such as those described by the author be effective?*

5. DISCRIMINATION LAW

Only some forms of unfairness are legally actionable. Can algorithms produce legally-prohibited discrimination?

Solon Barocas & Andrew Selbst, Big Data's Disparate Impact, California Law Review (2015), pp. 694-712

- *How well does algorithmic bias fit into the legal requirements for disparate impact?*

Jennifer Stark and Nicholas Diakopoulos, Uber Seems to Offer Better Service in Areas With More White People. That Raises Some Tough Questions, Washington Post, March 10, 2016

- *As the authors ask, "Should Uber be expected, or required even, to address this type of racial disparity in neighborhood service quality?"*

6. BUSINESS ANALYTICS ON TRIAL

Based on a “ripped from the headlines” episode of a TV drama, we’ll act out a realistic scenario of alleged algorithmic discrimination

Watch the “Good Wife” video segment on the Canvas site.

Prepare to assume your pre-assigned role in a mock trial.

7. DATA COLLECTION AND AGGREGATION

Are there limits on how data should be collected, used, and shared?

Neil Richards & Jonathan King, Big Data and the Future for Privacy (2016), pp. 3-15

- *What is privacy? Why is it important?*
- *How does big data change the way we think about privacy?*

Solon Barocas and Helen Nissenbaum, Big Data’s End Run Around Procedural Privacy Protections, Communications of the ACM (November 2014)

- *What do the authors believe that transparency and consent are insufficient?*

Kashmir Hill and Surya Mattu, How a Company You’ve Never Heard of Sends You Letters about Your Medical Condition, Gizmodo, June 19, 2017

Mara Hvistendahl, Inside China’s Vast Experiment in Social Ranking, Wired, December 14, 2017

8. PRIVACY LAW

Privacy is the subject of many legal and regulatory regimes in the U.S. and elsewhere. How well do those rules apply to big data and business analytics?

The White House, Big Data: Seizing Opportunities, Preserving Values (2014), pp. 15-21

- *What are the key elements of the U.S. approach to privacy law?*
- *Do you think the U.S. legal framework is effective in general? Will it be effective for the novel challenges of big data and business analytics?*

Paul M. Schwartz, Data Protection Law and the Ethical Use of Analytics, Centre for Information Policy Leadership (2010), pp. 18-26

- *How does the European approach to privacy differ from the U.S.?*
- *How should companies respond to the legal requirements around privacy?*

Matt Burgess, What is GDPR?, Wired UK, January 2, 2018

- *What new requirements does GDPR add?*
- *How would they be applied to algorithmic decision-making based on analytics?*

Kiel Brennan-Marquez, The Supreme Court’s Big Data Problem, Points, June 29, 2016

- *How is the growth of business analytics changing the legal boundaries of privacy?*

9. PERILS OF PREDICTION

If sensitive attributes can be inferred from other data, does it even make sense to talk about privacy any more?

Charles Duhigg, How Companies Learn Your Secrets, N.Y. Times Magazine, Feb. 16, 2012

- *How does Target analyze customer data to make inferences about customers?*
- *In your opinion, is the Target system an intrusion on privacy? Why or why not?*
- *Do Target's actions violate any legal rules?*
- *Do Target's actions violate any ethical norms?*
- *Should Target do anything differently?*

Michal Kosinski et al, Private Traits and Attributes are Predictable from Digital Records of Human Behavior, Proceedings of the National Academy of Sciences, April 9, 2013

- *What kinds of information can be predicted based on Facebook Likes?*
- *What could possibly go wrong?*

10. INFLUENCING USERS

To what extent does analysis itself influence behavior? And what are the limits on using analytics not merely to understand and predict customer actions, but to shape them?

Zeynep Tufekci, Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency, J. on Telecomm. & High Tech L. (2015), pp. 203-209

- *What was Facebook trying to achieve in its emotional contagion study?*
- *Why were Facebook's actions controversial?*
- *What is "algorithmic gatekeeping"? Why does Tufekci believe it is a concern?*

Tristan Harris, How Technology Hijacks People's Minds—from a Magician and Google's Design Ethicist, Medium, May 18, 2016

- *Why are the techniques that Harris describes effective?*
- *Do you share Harris' concerns about the addictive character of such approaches?*

Rebecca Rosen, Is This the Grossest Advertising Strategy of All Time?, The Atlantic, October 3, 2013

- *What exactly does the author find new and objectionable about this marketing approach?*

11. MARKET POWER AND INEQUALITY

Should we be concerned about algorithmic monopolies or other anti-competitive practices?

Jerry Useem, How Online Shopping Makes Suckers of Us All, The Atlantic Monthly, May 2017

- *Do you find algorithmic pricing practices troubling?*

Ryan Calo, Digital Market Manipulation, 82 George Washington Law Review 995 (2014), pp. 1003-1012, 1020-24

- *How do online intermediaries, in Calo's account, engage in forms of manipulation?*

Jonas Lerman, [Big Data and its Exclusions](#), 66 Stanford Law Review Online 55, September 3, 2013

- *How are the author's concerns about digital exclusion different from, and similar to, the issues of digital discrimination?*

Natalia Drozdiak, [EU Asks: Does Control of 'Big Data' Kill Competition?](#), Wall Street Journal, January 2, 2018

- *Why is the European Union worried that big data will reduce competition?*

12. STAKEHOLDER NEGOTIATION

You will be assigned a role in a group negotiation to address concerns about the dangers of unchecked use of business analytics.

Submit pre-negotiation worksheet for your role.

13. RESPONSES: TRANSPARENCY AND PROCESS

Proposals to address the kinds of concerns raised in this class generally fall into two categories: making algorithms transparent or interpretable, and ensuring procedural opportunities to review and challenge algorithmic decisions.

Cliff Kuang, [Can AI Be Taught to Explain Itself?](#), New York Times Magazine, Nov. 21, 2017

- *What are the benefits of interpretable algorithms?*
- *What are the difficulties and limitations of algorithmic transparency?*

Julia Powles, [New York City's Bold, Flawed Attempt to Make Algorithms Accountable](#), The New Yorker, December 20, 2017

- *Would the proposed New York algorithmic disclosure mandate be beneficial?*
- *Should there be similar disclosure when users are private companies or individuals?*

Neil Richards & Jonathan King, [Big Data and the Future for Privacy](#), Handbook of Research on Digital Transformations (2016), pp. 16-21

- *What is "soft regulation" and how is it different from ordinary regulation?*

Kate Crawford and Jason Schultz, [Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms](#), 55 Boston College Law Review pp. 121-28 (2014)

- *How is the "due process" approach different from existing laws we've discussed?*
- *Would this approach be effective?*

Houston Federation of Teachers v. Houston Independent School District (May 4, 2017)

- *On what grounds did the plaintiffs challenge the use of algorithmic "value added models" to evaluate public school teachers?*
- *Why did the plaintiffs win on the procedural due process claim? Why did they lose on the substantive due process claim?*

14. EXAM AND COURSE WRAP-UP

The closed-book examination will be done in class, and will cover all prior sessions.