The Good, The Bad, and The Creepy: Why Data Scientists Need to Understand Ethics

Jennifer Lewis Priestley, Ph.D.
Kennesaw State University
Dr. Priestley is the Associate Dean of The Graduate College and the Director of the Analytics and Data Science Institute at Kennesaw State University. In 2012, the SAS Institute recognized Dr. Priestley as the 2012 Distinguished Statistics Professor of the Year. She served as the 2012 and 2015 Co-Chair of the National Analytics Conference. Datanami recognized Dr. Priestley as one of the top 12 “Data Scientists to Watch in 2016.” She is a member of the Advisory Board for the Southern Data Science Conference.

Prior to receiving a Ph.D. in Statistics, Dr. Priestley worked in the Financial Services industry for 11 years. Her positions included Vice President of Business Development for VISA EU in London, as well as for MasterCard US and an analytical consultant with Accenture’s strategic services group. Dr. Priestley received a Ph.D. from Georgia State, a MBA from The Pennsylvania State University, and a BS from Georgia Tech.
Why Data Scientists Need to Understand Ethics

Issue 1: A few people can cause a great deal of harm

Major Data Breaches by Year

- Marriott
- Anthem
- Equifax
- Adult Friend Finder
- Ebay
- Yahoo

Individuals impacted (in millions)
Why Data Scientists Need to Understand Ethics

Issue 1: A few people can cause a great deal of harm

Should researchers utilize hacked datasets that have been released in public forums?
Why Data Scientists Need to Understand Ethics

Issue 2: Lack of Consent

Experimental evidence of massive-scale emotional contagion through social networks

Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock

PNAS June 17, 2014 111 (24) 8788-8790; first published June 2, 2014 https://doi.org/10.1073/pnas.1320040111

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved March 25, 2014 (received for review October 23, 2013)
DNA profiles from ancestry websites helped identify the Golden State Killer suspect

He wasn’t the first criminal to fall to familial DNA matching, and he won’t be the last.
Why Data Scientists Need to Understand Ethics

Issue 3: Will this algorithm do what I think it does?
Why Data Scientists Need to Understand Ethics

Issue 3: Will this algorithm do what I think it does?
Examples of Algorithmic Biases

- Pre-Existing
- Technical
- Emergent
Why Data Scientists Need to Understand Ethics

Issue 3: Will this algorithm do what I think it does?

Mission of the FDA:

The Food and Drug Administration is responsible for protecting the public health by ensuring the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices; and by ensuring the safety of our nation's food supply, cosmetics, and products that emit radiation.
What is the role of the academic community?

1) Ensure students understand the law (e.g., GDPR)
2) Create awareness...ask questions at each stage
3) Engage Institutional Research Boards
4) Don’t shortcut the math/statistics
Maslow’s Hierarchy of Data Science

- Mathematics
- Statistics, Computer Science, Programming
- Algorithm Development, Modeling, Classification
- Visualization, Storytelling, Communication
- Ethics

Disciplines from Science
Disciplines from Science and Business
Disciplines from Science, Business and the Humanities

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. © indicates USA registration. Other brand and product names are trademarks of their respective companies.
Thank you!

Contact Information
jpriestl@kennesaw.edu

Reminder:
Complete your session survey in the conference mobile app.