Ethical Considerations for Data Science Research
SURV 699X
1 credit/2 ECTS
Fall 2020

Instructor
Jessica Vitak, PhD, jvitak@umd.edu

Video lecture by Jessica Vitak, PhD

Short Course Description
The goal of research ethics is to protect human subjects from harm when they participate in a study. In the digital age, however, what constitutes "participation" has become blurry, especially with the rise of social media platforms and other online apps and services. Furthermore, new applications of big data raise important questions about how to protect consumers from harms, and what kinds of notice and consent should be obtained. This course provides and introduction and overview of research ethics in the 21st century and evaluates the many challenges to conducting ethical research.

Course and Learning Objectives
By the end of the course, students will...
- Describe the history of research ethics and the goals of institutional review boards
- Describe the challenges data science and big data raise for protecting individuals' rights and privacy
- Identify ethical issues in the study design, data collection, and data analysis process
- Detail best practices for conducting ethical research

Prerequisites
No prerequisites.

Class Structure and Course Concept
This is an online course using a flipped classroom design. It covers the same material and content as an on-site course but runs differently. In this course, you are responsible for watching video-recorded lectures and reading the required literature for each unit prior to participating in mandatory weekly one-hour online

meetings where students have the chance to discuss the materials from a unit with the instructor.

Although this is an online course where students have more freedom in when they engage with the course materials, students are expected to spend the same amount of time overall on all activities in the course – including preparatory activities (readings, studying), in-class-activities (watching prerecorded videos, attending the live online meetings), and follow-up activities (working on assignments) – as in an on-site course. As a rule of thumb, you can expect to spend approximately 3h/week on in-class-activities and 9 hours per week on out-of-class activities (preparing for class, readings, assignments, projects, studying for quizzes). Therefore, the workload in all courses will be approximately 12h/week. This is a 1-credit/2 ECTS course that runs for 4 weeks. Please note that the actual workload will depend on your personal knowledge.

Mandatory Weekly Online Meetings
Thursday, 12 PM EDT/6 PM CEST, starting September 24

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on https://www.elms.umd.edu/. If video participation via Internet is not possible, arrangements can be made for students to dial in and join the meetings via telephone.

In preparation for the weekly online meetings, students are expected to watch the lecture videos and read the assigned literature before the start of the meeting. In addition, students are required to post questions or comments about the materials covered in the videos or readings of the week in the forum before the meetings (deadlines for posting are Wednesday, 11:59 PM EDT/Thursday, 5:59 AM CEST).

Students have the opportunity to use the Zoom meeting room set up for this course to connect with peers outside the scheduled weekly online meetings (e.g., for study groups). Students are encouraged to post the times that they will be using the room to the course website forum to avoid scheduling conflicts. Students are not required to use Zoom and can use other online meeting platforms such as Google Hangout or Skype.

Grading
Grading will be based on:
- Participation in discussion during the weekly online meetings and contributions to weekly discussion forums demonstrating understanding of the required readings and video lectures (10% of grade)
- Four open-book quizzes assessing comprehension of course material (20% of grade; 5% each)

- Three online homework assignments reviewing specific aspects of the material covered (45% of grade; 15% each)
- Final paper covering overarching themes of the class (25%).

A+	100 - 97	C+	79-77	F	59 or below		
A	96 – 93	C	76-73				
A-	92 – 90	C-	72-70				
B+	89 – 87	D+	69-67				
B	86 – 83	D	66-63				
B-	82 – 80	D-	62-60				

The grading scale is a base scale recommended by the IPSDS. Variations for grading on a scale are at the discretion of the instructor.

Dates of when assignment will be due are indicated in the syllabus. Extensions will be granted sparingly and are at the instructor's discretion. If you know you will not be able to meet a deadline, email the professor before the due date to request an extension. If an assignment is submitted late and no extension has been given, a 10% penalty will be applied for each day it is late.

Technical Equipment Needs

The learning experience in this course will mainly rely on the online interaction between students and the instructor during the weekly online meetings. Therefore we encourage all students in this course to use a web camera and a headset. Decent quality headsets and web cams are available for less than $20 each. We ask students to refrain from using built-in web cams and speakers on their desktops or laptops. We know from our experience in previous online courses that this will reduce the quality of video and audio transmission and therefore will decrease the overall learning experience for all students in the course. In addition, we suggest that students use a wire connection (LAN), if available, when connecting to the online meetings. Wireless connections (WLAN) are usually less stable and might be dropped.

Long Course Description

Networked technologies—including the internet of things (IoT), wearables, ubiquitous sensing, social sharing platforms, and other AI-driven systems—are generating a tremendous amount of data about individuals, companies, and societies. These technologies provide a range of new opportunities for data scientists and researchers to understand human behavior and develop new tools that benefit society. At the same time, the ease with which data can be collected and analyzed raises a wide range of ethical questions about these technologies, their creators, and their users.

In recent years, we have seen numerous examples of research and technologies that are ethically problematic. For example, Facebook's Cambridge Analytica scandal revealed researchers using problematic tactics to collect profile data from millions of Facebook users. In addition, algorithms and machine learning techniques have been revealed as systematically biased in how they evaluate resumes[1], recommend parole for prisoners[2], decide where police units should deploy[3], and identify people through facial recognition technology[4], just to name a few.

Therefore, it is critical that data scientists and others who will be working with big data can critically assess the potential risks and benefits of any end products, whether they are developing a search engine or a tool for detecting terrorists. This course will provide an overview of key ethical issues that arise when working with big data, and it will provide opportunities to review and reflect on past mistakes in this space.

Readings

There is no text book for this course.

Required and recommended readings are provided below for each specific unit.

Academic Conduct

Clear definitions of the forms of academic misconduct, including cheating and plagiarism, as well as information about disciplinary sanctions for academic misconduct may be found at

https://www.president.umd.edu/sites/president.umd.edu/files/documents/policies/III-100A.pdf (University of Maryland) and

https://www.uni-mannheim.de/en/research/good-research-practice/ (University of Mannheim).

Knowledge of these rules is the responsibility of the student and ignorance of them does not excuse misconduct. The student is expected to be familiar with these guidelines before submitting any written work in this course. Lack of familiarity with these rules in no way constitutes an excuse for acts of misconduct. Charges of plagiarism and other forms of academic misconduct will be dealt with very seriously

---

[1] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G
[2] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
[3] https://www.newscientist.com/article/mg23631464-300-biased-policing-is-made-worse-by-errors-in-pre-crime-algorithms/
[4] https://www.nytimes.com/2019/04/03/technology/amazon-facial-recognition-technology.html

and may result in oral or written reprimands, a lower or failing grade on the assignment, a lower or failing grade for the course, suspension, and/or, in some cases, expulsion from the university.

Accommodations for Students with Disabilities

In order to receive services, students at the University of Maryland must contact the Accessibility & Disability Service (ADS) office to register in person for services. Please call the office to set up an appointment to register with an ADS counselor. Contact the ADS office at 301.314.7682; https://www.counseling.umd.edu/ads/.

Students at the University of Mannheim should contact the Commissioner and Counsellor for Disabled Students and Students with Chronic Illnesses at http://www.uni-mannheim.de/studienbueros/english/counselling/disabled_persons_and_persons_with_chronic_illnesses/.

Course Evaluation

In an effort to improve the learning experience for students in our online courses, students will be invited to participate in an online course evaluation at the end of the course (in addition to the standard university evaluation survey). Participation is entirely voluntary and highly appreciated.

Class Schedule

Please note that assignments and dates are subject to change. Information (e.g., articles and assignments) posted to the course website supersedes the information noted here.

Unit 1: Social and Ethical Implications of Data

Video lecture: available Thursday, September 17

Online meeting: Thursday, September 24, 12 PM EDT/6 PM CEST

Online quiz 1: due Wednesday, September 23, 11:59 PM EDT/Thursday, September 24, 5:59 AM CEST

Assignment 1: due Sunday, September 27, 5:59 PM EDT/11:59 PM CEST

Required readings:

Vitak, J., Shilton, K. & Ashktorab, Z. (2016). Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16). Association for Computing Machinery, New York, NY, USA, 941–953.

Boyd, D. & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society, 15(5), 662-679.

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. Big Data & Society, 3(1).

Recommended readings (optional):

Postman, N. (1998, March 28). Five Things We Need to Know About Technological Change.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems, 14(3), 330–347.

Moon, M. (2009). The History and Role of Institutional Review Boards: A Useful Tension. AMA Journal of Ethics.

By the end of this unit students will…
- Describe emerging ethical questions in the era of big data
- Identify key benefits and drawbacks of big data.
- Critique research that collects and analyzes "public" data.

Unit 2: Foundations of Research Ethics

Video lecture: available Thursday, September 24

Online meeting: Thursday, October 1st, 12 PM EDT/6 PM CEST

Online quiz 2: due Wednesday, September 30, 11:59 PM EDT/Thursday, October 1st, 5:59 AM CEST

Assignment 2: due Sunday, October 4, 5:59 PM EDT/11:59 PM CEST

Required readings:

Metcalf, J. (2014). "Ethics Codes: History, Context, and Challenges." Council for Big Data, Ethics, and Society.

Velasquez, M., Andre, C., Shanks, S.J., T., & Meyer, M.J. "What is ethics?" Center for Applied Ethics, Santa Clara University. https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/what-is-ethics/

Jackman, M. & Kanerva, L. (2016). Evolving the IRB: Building Robust Review for Industry Research. 72 Wash. & Lee L. Rev. Online 442.

Association of Internet Researchers (AoIR) Code of Ethics: https://aoir.org/reports/ethics3.pdf

Recommended readings (optional):

Saltz, J. S., & Dewar, N. (2019). Data science ethical considerations: a systematic literature review and proposed project framework. https://www.semanticscholar.org/paper/Data-science-ethical-considerations%3A-a-systematic-Saltz-Dewar/6f26eca39167e40f0ac5484442b41db8b9aa1e92

Tiell, S., & Metcalf, J. (2016). The Universal Principles of Data Science Ethics. Accenture Labs.

By the end of this unit, students will…

- Describe key ethical questions for research using big data.
- Identify and describe concepts important to codes of ethics.
- Critique existing codes of ethics regarding their treatment of big data.
- Identify ethical problems with a study design using big data.

Unit 3: Ethical Concerns Around Data Collection, Storage, and Analysis

Video lecture: available Thursday, October 1st

Online meeting: Thursday, October 8, 12 PM EDT/6 PM CEST

Online quiz 3: due Wednesday, October 7, 11:59 PM EDT/Thursday, October 8, 5:59 AM CEST

Assignment 3: due Sunday, October 11, 5:59 PM EDT/11:59 PM CEST

Required readings:

Zimmer, M. (2010). "But the data is already public": On the ethics of research in Facebook. Ethics and Information Technology, 12(4), 313–325.

Barocas, S., & Nissenbaum, H. (2014, November). Big data's end run around procedural privacy protections: Recognizing the inherent limitations of consent and anonymity. Communications of the ACM, 57(11), 31-33.

Ohm, P. (2009). Broken promises of privacy: Responding to the surprising failure of anonymization. UCLA Law Review, 57, 1701. (read pages 1701-1731)

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Recommended readings (optional):

Zimmer, M. (2016, May 14). OkCupid Study Reveals the Perils of Big-Data Science. Wired. https://www.wired.com/2016/05/okcupid-study-reveals-perils-big-data-science/

Vetro, A., Santangelo, A., Beretta, E. and De Martin, J.C. (2019). AI: from rational agents to socially responsible agents. Digital Policy, Regulation and Governance, 21, 291-304. https://doi.org/10.1108/DPRG-08-2018-0049

By the end of this unit, students will…
- Identify new concerns that have arisen for research involving digital data/big data.
- Design a study design addresses ethical concerns around consent, anonymization, and public data.
- Describe the pros and cons of aggregating multiple data sources.
- Describe best practices for storing and analyzing data.

Unit 4: Best Practices for Conducting Ethical Research

Video lecture: available Thursday, October 8

Online meeting: Thursday, October 15, 12 PM EDT/6 PM CEST

Online quiz 4: due Wednesday, October 14, 11:59 PM EDT/Thursday, October 15, 5:59 AM CEST

Required readings:

Tiell, S. & Metcalf, J. (2016). The Ethics of Data Sharing: A guide to best practices and governance. Accenture.

Keyes, O. (2019, April 8). Counting the Countless: Why data science is a profound threat for queer people. Real Life.

Olteanu, A., Castillo, C., Diaz, F., & Kiciman, E. (2019). Social data: Biases, methodological pitfalls, and ethical boundaries. Front. Big Data 2:13.

Barocas, S., & Boyd, D. (2017). Engaging the ethics of data science in practice. Communications of the ACM, 60(11), 23-25.

Recommended readings (optional):

Makhoul, J., Chehab, R. F., Shaito, Z., & Sibai, A. M. (2018). A scoping review of reporting 'Ethical Research Practices' in research conducted among refugees and war-affected populations in the Arab world. BMC medical ethics, 19(1), 36.

By the end of this unit, students will...
- Create and implement an ethical research checklist to apply to new studies
- Critically evaluate research studies regarding their potential for social harm
- Articulate ethical challenges to conducting research with marginalized groups

Final Paper

Due: Thursday, October 22nd, 12 PM EDT/6 PM CEST