

Syllabus

Ethical Considerations for Data Science Research

1 credit/2 ECTS

Jessica Vitak

Video lecture by

Jessica Vitak

July 17-August 7, 2023

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 an introduction and overview of research ethics in the 21st century and evaluates the many challenges to conducting ethical research.

Course 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

None.

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

Mondays, 12PM-1PM ET / 6PM-7PM CET, starting July 17, 2023

Meetings will be held online through Zoom. Follow the link to the meeting sessions on the course website on mannheim.instructure.com. 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 should post questions and discussion comments about the materials covered in the videos and readings of the week in the forum before the meetings (deadline for posting questions is Wednesdays 5PM ET/11PM CET).

Students have the opportunity to use the Conferences feature in Canvas to connect with peers outside the scheduled weekly online meetings (e.g., for study groups). Students are not required to use Canvas Conferences and can of course use other online meeting platforms such as Google Hangouts, Skype or Microsoft Teams.

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 (15% 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 (20%).

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 MDM. Variations for grading on a scale are at the discretion of the instructor.

The final grade will be communicated under the assignment "Final Grade" in the Canvas course. Please note that the letter grade written in parentheses in Canvas is the correct final grade. The point-grade displayed alongside the letter grade is irrelevant and can be ignored. 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 the students and the instructors 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.

Mannheim Business School would also like to officially inform you that, in order to facilitate your participation in this course, your personal data will be processed by and on systems run by MBS and our subcontractors. You can find detailed information in our privacy policy and information for data subjects [here](#).

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 resumes¹, recommend parole for prisoners², decide where police units should deploy³, and identify people through facial recognition technology⁴, 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

¹ <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrapes-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>

There is no textbook 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 or taking any exams 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 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 Mannheim Business School 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. Participation is entirely voluntary and highly appreciated.

Sessions

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.

Week 1:

Video lecture available: Monday, July 10, 2023

Online meeting: Monday, July 17, 2023, 12PM ET/6PM CET

Online quiz 1 due: Sunday, July 16, 2023, 11:59PM ET / Monday, July 17, 2023, 5:59AM CET

Assignment 1 due: Thursday, July 20, 2023, 5:59PM EDT / 11:59PM CEST

Required readings:

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.

<https://doi.org/10.1080/1369118X.2012.678878>

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951716650211>

Moon, M. (2009). The history and role of Institutional Review Boards: A useful tension. *AMA Journal of Ethics*, 11(4), 311-316. <https://doi.org/10.1001/virtualmentor.2009.11.4.pfor1-0904>

Recommended readings (optional):

Postman, N. (1998, March 28). Five things we need to know about technological change. *Talk delivered in Denver, Colorado*. <https://web.cs.ucdavis.edu/~rogaway/classes/188/materials/postman.pdf>

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330-347. <https://dl.acm.org/doi/10.1145/230538.230561>

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.

Week 2:

Video lecture available: Monday, July 17, 2023

Online meeting: Monday, July 24, 2023, 12PM ET/6PM CET

Online quiz 2 due: Sunday, July 23, 2023, 11:59PM ET / Monday, July 24, 2023, 5:59AM CET

Assignment 2 due: Thursday, July 27, 2023, 5:59PM EDT / 11:59PM CEST

Required readings:

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/>

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>

Metcalfe, J. (2014). "Ethics codes: History, context, and challenges." Council for Big Data, Ethics, and Society. <https://bdes.datasociety.net/council-output/ethics-codes-history-context-and-challenges/>

Recommended readings (optional):

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

Vitak, J., Shilton, K., & Ashktorab, Z. (2016). Beyond the Belmont Principles: Ethical challenges, practices, and beliefs in the online data research community. Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW) (pp. 941-953). New York: ACM. <https://dl.acm.org/doi/10.1145/2818048.2820078>

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.

Week 3:

Video lecture available: Monday, July 24, 2023

Online meeting: Monday, June 31, 2023, 12PM ET/6PM CET

Online quiz 3 due: Sunday, July 30, 2023, 11:59PM ET / Monday, June 31, 2023, 5:59AM CET

Assignment 3 due: Thursday, August 3, 2023, 5:59PM EDT / 11:59PM 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.

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):

Ohm, P. (2009). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57, 1701. <https://www.uclalawreview.org/pdf/57-6-3.pdf>

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.

Week 4:

Video lecture available: Monday, June 31, 2023

Online meeting: Monday, August 7, 2023, 12PM ET/6PM CET

Online quiz 4 due: Sunday, August 6, 2023, 11:59PM ET / Monday, August 7, 2023, 5:59AM CET

Required readings:

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

Olteanu, A., Castillo, C., Diaz, F., & Kiciman, E. (2016). Social data: Biases, methodological pitfalls, and ethical boundaries. *Frontiers in Big Data*, 2, Article 13. <https://doi.org/10.3389/fdata.2019.00013>

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

Sloane, M. (2022). To make AI fair, here's what we must learn to do. *Nature*, 605, 9.
<https://www.nature.com/articles/d41586-022-01202-3>

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.

Keyes, O. (2019, April 8). Counting the Countless: Why data science is a profound threat for queer people. *Real Life*. <https://reallifemag.com/counting-the-countless/>

Kearns, M., & Roth, A. (2020). Ethical algorithm design should guide technology regulation. Brookings Institution. <https://www.brookings.edu/research/ethical-algorithm-design-should-guide-technology-regulation/>

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

Final Paper Due: Sunday, August 13, 2023, 12PM ET/6 PM CET