

Spring, 2024 – CDS DS380: Data, Society, and Ethics

Course Time & Location: lectures and case studies are Tue, Thu 11:00am-12:15pm in KCB 106, 565 Commonwealth Avenue; discussion sections:

- Wed 12:20pm-1:10pm in CDS 164, with Jasmine
- Wed 1:25-2:15 in CDS 164, with Carmen

Course Credits: 4

Instructor Information:

Instructor Name: Wesley J. Wildman

Office Location: CDS 1641

Contact Information: wwildman@bu.edu

Office Hours: Tuesdays 2:00-3:00 in CDS 1641, Wed 10:00-11:00 in STH 335, and by appointment

Teaching Assistant Information:

- Carmen Pelayo Fernandez (cpelayo@bu.edu), office hours Wednesdays 2:20pm-3:20pm, CDS 16th floor bullpen
- Jasmine Pham (jazzmine@bu.edu), office hours Fridays 1:00pm-2:00pm, CDS 16th floor bullpen

Course Description

A society overflowing with data demands computational and data-science methods to interpret and leverage that wealth of information. Collecting, analyzing, and communicating data inevitably provokes profound ethical challenges, and they will only multiply in number and complexity as AI methods for handling data continue to develop. Welcome to Cyber-ethics. AI methods are here to stay; how do we achieve an accurate assessment of the utility and dangers of AI methods, avoiding the extremes of misplaced confidence and undue fear? Possession of data often confers power over others; should this be accepted or resisted or regulated? Bias plagues datasets and the computer algorithms used to support their interpretation; how do we detect and correct bias to avoid amplifying existing social inequities and creating new forms of unfairness? Data is being collected and shared all the time; what does data privacy mean and to what extent should it be an individual right? In an intensification of older methods of social control, AI techniques are routinely being applied to data for manipulating public perception and opinion in advertising and politics; how do we determine where to draw a line between freedom of expression and unacceptable manipulation? And how does the growing struggle over cyber-ethics interact with interpretations of the human condition and social justice upheld by wisdom traditions for centuries? This course addresses such questions by means of a series of real-world case studies presented alongside ethical tools and analyses that are useful both for staying alert to emerging ethical challenges and responding to them as they arise in both employment settings and everyday life. Students will complete several exercises during the

course to demonstrate their facility with key ethics tools and techniques and will develop a detailed ethical analysis of one case study as a final project. Students will meet in discussion sections to analyze all of these issues in more detail.

CDS Learning Outcomes

Learning Outcome 1: *Students will learn (1) to discern morally charged issues in novel situations, and (2) to reason persuasively about such issues based on ethical first principles and relevant professional codes of ethics.*

Learning Outcome 2: *Students will learn to deal with interpersonally and socially charged issues, such as gender and racial inequality or cultural parochialism and paternalism, and to engage diverse stakeholders and fellow professionals respectfully and productively.*

Learning Outcome 3: *Students will learn how to handle ethical challenges related to computational and data sciences as they arise in the workplace and in their own personal lives.*

Hub Learning Outcomes

This course intends to satisfy requirements for three hub learning goals.

Diversity, Civic Engagement, and Global Citizenship: Ethical Reasoning

Learning Outcome I: *Students will be able to identify, grapple with, and make a judgment about the ethical questions at stake in at least one major contemporary public debate, and engage in a civil discussion about it with those who hold views different from their own.*

The explosion of big data, machine learning, and other computational and data-science technologies in recent decades has introduced a host of novel ethical challenges. In this course, students will be introduced to fundamental principles of ethical reasoning and will examine their applicability to a variety of vexing ethical challenges related to digital technologies, including algorithmic bias, data monetization and surveillance capitalism, disinformation and political disenfranchisement, and the protection of privacy amidst panoptic data gathering. These ethical challenges will be addressed in lectures, explored through independent student research, and discussed with civility and respect in formal in-class case studies and debates.

Learning Outcome II: *Students will demonstrate the skills and vocabulary needed to reflect on the ethical responsibilities that face individuals (or organizations, or societies or governments) as they grapple with issues affecting both the communities to which they belong and those identified as “other.” They should consider their responsibilities to future generations of humankind, and to stewardship of the Earth.*

The stakes for addressing the ethical challenges brought on by novel computational and data-science technologies could not be higher: our ability to address collectively the growing climate crisis, to safeguard representative forms of government, and to continue expanding human rights across the globe all critically depend on ensuring the ethical use of digital data and analytics technologies. Students will gain a deep understanding of some of the most pressing challenges in digital ethics as well as a clear sense of the ethical responsibilities of the creators and users of those technologies. They will also gain an appreciation for the inherent limitations

of governments and other authorities when it comes to regulating novel technologies and the corresponding importance of individual stakeholders adopting high standards of ethical conduct.

Scientific and Social Inquiry: Social Inquiry II

Note: BU HUB's "Social Inquiry I" is a prerequisite

Learning Outcome I: *Students will apply principles and methods from the social sciences based on collecting new or analyzing existing data in order to address questions, solve problems, or deepen understanding. They will understand the nature of evidence employed in the social sciences and will demonstrate a capacity to differentiate competing claims in such fields. This includes reflecting on and critically evaluating how social scientists formulate hypotheses, gather empirical evidence of multiple sorts, and analyze and interpret this evidence.*

Throughout this course, students will examine a host of social problems related to AI and big data, each problem being presented in a case-study format accompanied by relevant real-world data. Students will gain an appreciation for the importance of following ethical guidelines governing the collection, processing, interpretation, and dissemination of social data and the correlated risks of unethical behavior during these stages of research.

Learning Outcome II: *Using their knowledge of the natural and social sciences, students will engage with issues of public policy, such as climate change, inequality, and health that involve the intersection of perspectives from different disciplines. This would entail an ability to identify the evidentiary basis for scientific claims, the challenges to it, and the connections among the economic, social, and scientific factors that shape the creation and adoption of effective public policy.*

Students will gain an appreciation of the complexities of developing effective public policy through their final project, which requires them to advocate for a specific policy aimed at preventing or mitigating one of the unethical uses of data and AI explored during the course. Through in-depth research, students will gain a sophisticated understanding of the problem in question. They will also need to anticipate and respond to criticisms of their policy proposal, as well as consider whether their proposed policy might lead to unintended negative consequences.

Intellectual Toolkit: Research and Information Literacy

Learning Outcome I: *Students will be able to search for, select, and use a range of publicly available and discipline-specific information sources ethically and strategically to address research questions.*

For each weekly case study, the class will ask about the quality of information informing the description and analysis of the ethical problem being examined. Students will develop criteria for detecting reliable versus unreliable data, and learn a variety of methods for employing publicly available data for research purposes.

Learning Outcome II: *Students will demonstrate understanding of the overall research process and its component parts, and be able to formulate good research questions or hypotheses, gather and analyze information, and critique, interpret, and communicate findings.*

Students will be introduced to several models of the research process, including methods relevant to the ethical analysis of digital technologies. They will explicitly adopt one such model for their final projects, clearly formulating a policy-related research question, an hypothesis in the form of a recommended policy, and a body of information relevant to assessing that policy proposal. They will demonstrate skills in self-critique, sophisticated interpretation, and clear communication of their policy proposal and supporting argumentation.

Instructional Format, Course Pedagogy, and Approach to Learning

This class gathers twice each week for 75 minutes each meeting.

One class meeting presents lecture content with Q&A. The aim of these interactive lectures is to cultivate within students a cyber-ethics toolkit that is directly relevant to their personal lives and professional careers. This toolkit features fundamental ethics principles, relevant professional codes of ethics, social-science methods that orient cyber-ethics, data-analysis methods that support ethical analysis, logical reasoning for ethics, tendencies to cognitive error that impact ethical analysis, and descriptions of computational and data-science technologies necessary for relevant ethical insight. The two short paper assignments for the class pick up on key elements of lectures: professional codes of ethics and methods of detecting bias in data, analysis, or reasoning. The mid-term and final exams reinforce the lecture material.

The other class meeting each week presents a digital-ethics case study. The aim of these case studies is to bring real-world storytelling into the classroom so that students can take memorable and practical insights out of the classroom. Each case will focus on a recent, ethically problematic, real-world situation involving the use of data or data-analytics methods. Each case tells a fascinating story with thought provoking characters and dramatic twists, often provoking conflicting ethical perspectives while demanding ethical decisions. Students will empathize with characters in the case to feel the real weight of potential ethical or policy decisions. The case studies will employ a variety of formats, including problem-oriented discussion, public hearing, and formal debate. The final project for this class, a policy- focused ethical analysis paper, can be modeled on these case studies.

In addition, students gather with the teaching staff in a section meeting to discuss course content in greater detail.

Books and Other Course Materials

There is a textbook, available through Boston University's Barnes & Noble Bookstore:

Tavani, H. T. *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing*, 5th edition (Hoboken, NJ: Wiley, 2016). See [here](#) for information about this book.

Courseware

The course's website is available to registered students through Boston University's Blackboard Learn tool. Access through <https://onlinecampus.bu.edu>.

Assignments and Grading

Personalized professional code of ethics paper (1,000 words)

This paper will help students integrate and internalize the foundational ethical concepts presented in Part 1 of the course (weeks 1-3). Students will briefly explain their ethical philosophy, drawing on material presented in class and in the textbook, and then construct their own professional code of ethics, fitted to their anticipated career trajectory. This professional code of ethics should include 5-10 principles, constructed in dialogue with the principles advanced in professional codes of ethics such as those put forward by the Association for Computing Machinery or the Institute of Electrical and Electronics Engineers and discussed in class. The description of each principle, the explanation of its importance, and the justification of its inclusion must be in the student's own words, and should connect to the student's own life circumstances and specific career plans. The paper is to be submitted via SafeAssign on the course's Blackboard site to protect against plagiarism. Students should run their papers through GPTZero.me and originality-checking tools before submission, in accordance with the Generative AI Assistance Policy, to avoid false-positive reports; keep in mind that graders will also do this. Grader comments will be available on the site within one week of the deadline. Late papers will not be graded.

Detecting and correcting bias paper (1,000 words)

This paper will help students integrate and internalize the material on bias and amplifying inequities examined in Part 2 of the course (weeks 4-6). Students will develop a case study similar to the case studies presented in class. The case study will focus on a situation of alleged bias in a machine-learning algorithm, taking account of the underlying training dataset. The paper will describe the alleged problem, supply an explanation for the supposed problem, and propose a solution, either to the problem or to the public perception of a problem (whichever is relevant). Students may optionally statistically analyze the training dataset (this is NOT required). Students will specify the ethical principles employed in their analysis and contemplate how both the alleged problem and their proposed solution might impact public policy. The paper is to be submitted via SafeAssign on the course's Blackboard site to protect against plagiarism. Students should run their papers through GPTZero.me and originality-checking tools before submission, in accordance with the Generative AI Assistance Policy, to avoid false-positive reports; keep in mind that graders will also do this. Grader comments will be available on the site within one week of the deadline. Late papers will not be graded.

Final project: policy paper (2,000 words)

Students will write a white paper advocating a policy, or several related policies, aimed at addressing a particular problem in digital ethics. The paper must be based on extensive research of the associated ethical problem and must cite a minimum of a dozen appropriate

sources. It must also clearly delineate the relevant stakeholders and explain the ethical principles at issue. The paper must advocate for one or more policies intended to help address the problem and should specify the level at which the policy might be implemented (international organizations, national governments, associations of data scientists, data science watchdogs within private corporations, etc.). Finally, the white paper should anticipate and respond to likely critiques of the policy/policies and consider whether its implementation might lead to any unintended negative consequences. The paper is to be submitted via SafeAssign on the course's Blackboard site to protect against plagiarism. Students should run their papers through GPTZero.me and originality-checking tools before submission, in accordance with the Generative AI Assistance Policy, to avoid false-positive reports; keep in mind that graders will also do this. Grader comments will be available on the site within one week of the deadline. Late papers will not be graded.

Midterm and final exams

Both exams will consist of questions that will be machine-graded. They will test students' grasp of the ethical principles and key concepts explored during the course and in the assigned readings, and through the case studies. The midterm exam covers Part 1 and Part 2 of the course; the final exam covers Part 3 and Part 4 of the course.

Attendance, preparation, and participation

This refers to three areas: physical attendance, pre-class preparation, and in-class participation.

Goal setting and evaluation

Following the method of sociologist Alanna Gillis ("Reconceptualizing participation grading as skill building," *Teaching Sociology*, 47(1) 2019: 10-21), students will set one concrete goal (with a corresponding plan) for each of the three areas of Attendance, Preparation, and Participation, submitting this on Blackboard at the start of the course. Mid-way and at the end of the semester, students will assess their progress toward their goals, adding a sentence or two of justification. This serves as part of the basis for the Attendance, Preparation, and Participation component of the final grade.

Evaluating attendance

In-person attendance at class meetings is required. A good classroom experience requires a collaborative effort, particularly during case studies, and a student's absence detracts from the learning process of the whole. More than two unexcused absences will result in a demarcation of the final class grade (e.g. an 'A' will be reduced to an 'A-' – not merely for the Attendance, Preparation, and Participation component of the final grade but for the overall grade itself).

Evaluating pre-class preparation and in-class participation

After each class meeting, students will employ an adaptation of a rubric from J. R. Howard (*Discussion in the College Classroom*, San Francisco: Jossey-Bass, 2015) and report their score on Blackboard: 1=I did not complete the assigned reading and I did not participate in discussion, 2=I did not complete the assigned reading but I did contribute once, 3=I completed the assigned reading but did not contribute, 4=I completed the assigned reading and contributed more than

once, and 5=I completed the assigned reading, contributed more than once, and demonstrated listening skills by building on the comments of others and by encouraging others to contribute.

The teaching staff will stake these self-evaluations into account when assigning a score for the Attendance, Preparation, and Participation component of the final grade.

The real meaning of presence

Presence means in part being physically present in the classroom, appropriately rested, read-up, and ready to work. The teaching team's expectation is that you will come to class prepared to discuss the assigned readings and engage lecture material and case studies. Discussion involves speaking, active listening, sharing time and attention, asking others for clarification, and raising questions that move the conversation towards deeper insights.

Presence also means being attentive. Students are expected to use computing devices intelligently and creatively. This can mean going on to the web during class to investigate a concept or to dig up new perspectives on a case study. But students checking social media or news sites or streaming media during class are misunderstanding the point of university education and should remove themselves from the classroom. Attention is a precious and endangered practice, and we believe that it is an essential practice for deep intellectual engagement and effective learning. So we ask you, for the time you are in the class, to be fully present.

Grading breakdown

1. Personalized professional code of ethics paper (10%)
2. Detecting and correcting bias paper (15%)
3. Final project: policy analysis (25%)
4. Midterm exam (20%)
5. Final exam (20%)
6. Attendance, preparation, and participation (10%)

Resources/Support/How to Succeed in This Course

The teaching staff is dedicated to helping you succeed in this course. To that end, students should make use of office hours with the teaching staff and tutoring and writing services within the university.

If you are a student with a disability or believe you might have a disability that requires accommodations, please contact the Office for Disability Services (ODS) at (617) 353-3658 or access@bu.edu to coordinate any reasonable accommodation requests. ODS is located at 25 Buick Street on the 3rd floor.

Outline of Class Meetings: Date, Topic, Readings Due, Assignments Due

Part 1: Approaching Digital Ethics

Week 1: Introduction

Thu Jan 18

Interactive Lecture: introducing the course and motivating learning goals by laying out some of the ethical challenges of emerging computational and data-science technologies

Required reading:

- Generative AI Assistance Policy <https://www.bu.edu/cds-faculty/culture-community/gaia-policy/>

Tue Jan 23

Case Study: wrongful arrest based on facial recognition software

Required reading: at least two of the following:

- “‘The Computer Got It Wrong’: How Facial Recognition Led To False Arrest Of Black Man” <https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig>
- “Man Wrongfully Arrested Due to Facial Recognition Software Talks about “humiliating” Experience” <https://www.nbcnews.com/business/business-news/man-wrongfully-arrested-due-facial-recognition-software-talks-about-humiliating-n1232184>
- “Facial recognition linked to a second wrongful arrest by Detroit police” <https://www.engadget.com/facial-recognition-false-match-wrongful-arrest-224053761.html>
- “Facial recognition leads to first wrongful U.S. arrest, activists say” <https://www.nbcnews.com/tech/security/facial-recognition-leads-first-wrongful-u-s-arrests-activists-say-n1231971>
- “A black man was wrongfully arrested because of facial recognition” <https://www.theverge.com/2020/6/24/21301759/facial-recognition-detroit-police-wrongful-arrest-robert-williams-artificial-intelligence>
- “Wrongfully Accused by an Algorithm” <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>

Week 2: Misplaced Confidence in AI or a bright future?

Thu Jan 25

Interactive Lecture: ethical reasoning about the future

Required reading:

- Tavani ch. 1: “Introduction to Cyberethics: Concepts, Perspectives, and Methodological Frameworks”
- Max Tegmark, “Myths and Facts about Superintelligent AI” <https://www.youtube.com/watch?v=3Om9ssTm194>
- Janelle Shane, “The Danger of AI is Weirder than You Think” <https://www.bing.com/videos/search?q=danger+of+ai&&view=detail&mid=560DD092D783B9A06AA6560DD092D783B9A06AA6&>
- Kelsey Piper, “The case for taking AI seriously as a threat to humanity” <https://www.vox.com/future-perfect/2018/12/21/18126576/ai-artificial-intelligence-machine-learning-safety-alignment>

Tue Jan 30

Case Study: Uber car hits and kills woman in Arizona

Required reading: at least two of the following:

- “Self-driving Uber car that hit and killed woman did not recognize that pedestrians jaywalk” <https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281>
- “Death of Elaine Herzberg” https://en.wikipedia.org/w/index.php?title=Death_of_Elaine_Herzberg&oldid=968453777
- “Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam” <https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html>
- “Who was really at fault in fatal Uber crash? Here's the whole story” <https://www.azcentral.com/story/news/local/tempe/2019/03/17/one-year-after-self-driving-uber-rafaela-vasquez-behind-wheel-crash-death-elaine-herzberg-tempe/1296676002/>
- In-depth story on driver: <https://niemanstoryboard.org/stories/investigative-journalism-public-records-narrative-structure-uber-fatal-crash/>
- “Rafaela Vasquez pleads guilty” <https://www.azcentral.com/story/news/local/tempe/2023/07/28/rafaela-vasquez-pleads-guilty-in-fatal-uber-self-driving-crash-killed-pedestrian-elaine-herzberg/70488361007/>
- Controversy over paying for the burial of Elaine Herzberg <https://www.dailymail.co.uk/news/article-5578793/Family-woman-killed-Ubers-driverless-car-say-afford-bury-her.html>

Week 3: Ethical Foundations and Principles of Digital Ethics

Thu Feb 1: ethics principles and frameworks borrowed from philosophy

Required reading:

- Tavani ch. 2: “Ethical Concepts and Ethical Theories: Frameworks for Analyzing Moral Issues”
- Tavani ch. 3: “Critical Reasoning Skills for Evaluating Disputes in Cyberethics”

Tue Feb 6

Case Study: professional codes of ethics for IEEE, ACM, American Statistics Association, Data Science Association, and the European Commission, and how to approach writing your own **personal** code of ethics.

Required reading:

- Tavani ch. 4: “Professional Ethics, Codes of Conduct, and Moral Responsibility”
- “The Pledge of the Computing Professional” <http://pledge-of-the-computing-professional.org/home-page/the-oath>

Required reading: at least two specific codes of ethics:

- “IEEE Code of Ethics” <https://www.ieee.org/about/corporate/governance/p7-8.html>
- “ACM Code of Ethics and Professional Conduct” <https://www.acm.org/code-of-ethics>
- “Ethical Guidelines for Statistical Practice” (American Statistics Association) <https://www.amstat.org/your-career/ethical-guidelines-for-statistical-practice>
- “Data Science Code of Professional Conduct” (Data Science Association) <https://www.datascienceassn.org/code-of-conduct.html>
- “Code of Conduct” (Academic Data Science Association) <https://academicdatascience.org/code-of-conduct>
- “Ethics guidelines for trustworthy AI” (European Commission) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- “Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People” (US Office of Science and Technology Policy) <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

Part 2: Amplifying Inequities

Week 4: Fairness

Thu Feb 8

Interactive Lecture: ethics of distributive justice and fairness

Required reading:

- Tavani ch. 10: “The Digital Divide, Democracy, and Work”

Tue Feb 13

Case Study: Apple card credit limit gender bias/Gender bias in Amazon recruiting tool

Required reading: for the Apple case study, at least two of the following:

- “The Apple Card Didn't 'See' Gender—and That's the Problem” <https://www.wired.com/story/the-apple-card-didnt-see-genderand-thats-the-problem/>
- “Apple’s credit card is being investigated for discriminating against women” <https://www.theverge.com/2019/11/11/20958953/apple-credit-card-gender-discrimination-algorithms-black-box-investigation>
- “Apple Card Investigated After Gender Discrimination Complaints” <https://www.nytimes.com/2019/11/10/business/Apples-credit-card-investigation.html>
- “Apple Card is accused of gender bias. Here's how that can happen” <https://www.cnn.com/2019/11/12/business/apple-card-gender-bias/index.html>

Required reading: for the Amazon case study, at least two of the following:

- “Amazon reportedly scraps internal AI recruiting tool that was biased against women” <https://www.theverge.com/2018/10/10/17958784/ai-recruiting-tool-bias-amazon-report>

- “Amazon’s sexist AI recruiting tool: how did it go so wrong?”
<https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e>
- “Why it's totally unsurprising that Amazon's recruitment AI was biased against women”
<https://www.businessinsider.com/amazon-ai-biased-against-women-no-surprise-sandra-wachter-2018-10>
- “Amazon built an AI tool to hire people but had to shut it down because it was discriminating against women” <https://www.businessinsider.com/amazon-built-ai-to-hire-people-discriminated-against-women-2018-10>

Wed Feb 14 at 11:59pm: Deadline: paper #1 (code of ethics)

Week 5: Social Goods

Thu Feb 15

Interactive Lecture: ethical reasoning about social goods

Required reading:

- Tavani ch. 11: “Online Communities, Virtual Reality, and Artificial Intelligence”

Tue Feb 20

Case Study: Arnold pretrial risk assessment tool/COMPAS recidivism assessment tool

Required reading: at least two of the following:

- “Pretrial Risk Assessment Now Available to All Interested Jurisdictions; Research Advisory Board Announced” <https://www.arnoldventures.org/newsroom/laura-and-john-arnold-foundation-makes-pretrial-risk-assessment-available-to-all-jurisdictions-announces-expert-panel-to-serve-as-pretrial-research-advisory-board/>
- “Predictive policing algorithms are racist. They need to be dismantled.”
<https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/>
- “New Data Suggests Risk Assessment Tools Have Little Impact on Pretrial Incarceration”
<https://theappeal.org/new-data-suggests-risk-assessment-tools-have-little-impact-on-pretrial-incarceration/>
- “Advancing Pretrial Policy and Research” <https://advancingpretrial.org/>
- “States vs. Vendors: Are Some Risk Assessment Tools Better Than Others?”
<https://www.law.com/legaltechnews/2020/07/14/states-vs-vendors-are-some-risk-assessment-tools-better-than-others/>
- “A Bail Reform Tool Intended to Curb Mass Incarceration Has Only Replicated Biases in the Criminal Justice System” <https://theintercept.com/2020/07/12/risk-assessment-tools-bail-reform/>

Week 6: Reasoning

Thu Feb 22

Interactive Lecture: fallacies in ethical reasoning

Required reading:

- Tavani ch. 3: "Critical Reasoning Skills for Evaluating Disputes in Cyberethics"
- Radford University, "Identifying Fallacious Reasoning" <https://viva.pressbooks.pub/phi220ethics/chapter/identifying-fallacious-reasoning/>
- Figmentums, "Ethical Fallacies" <https://figmentums.com/2016/10/14/ethical-fallacies/>

Tue Feb 27

Case Study: Nextdoor reduction in racial profiling/Racial bias in hate speech detector

Required reading: at least two of the following:

- "How Nextdoor reduced racist posts by 75%" <https://splinternews.com/how-nextdoor-reduced-racist-posts-by-75-1793861389>
- "Nextdoor Removes App's 'Forward to Police' Feature" <https://www.nytimes.com/2020/06/23/us/nextdoor-forward-to-police-.html>
- "Reducing Racial Profiling on Nextdoor" <https://blog.nextdoor.com/2016/08/24/reducing-racial-profiling-on-nextdoor/>
- "How Nextdoor Addressed Racial Profiling on Its Platform" <https://hbr.org/2018/05/how-nextdoor-addressed-racial-profiling-on-its-platform>

Interlude

Thu Feb 29: GenAI in Higher Education

Required reading:

Generative AI Assistance Policy <https://www.bu.edu/cds-faculty/culture-community/gaia-policy/>

Tue Mar 5: Midterm exam (in class; bring computers)

Part 3: Invasions of Privacy

Week 7: Data Privacy and Security

Thu Mar 7

Interactive Lecture: ethics of privacy and security

Required reading:

- Tavani ch. 5: "Privacy in Cyberspace"
- Tavani ch. 6: "Security and Cyberspace"

Tue Mar 12: NO CLASS, SPRING RECESS

Thu Mar 14: NO CLASS, SPRING RECESS

Tue Mar 19

Case Study: Equifax/Capital One breach of customers' data

Required reading: read two of the following

- “EPIC - Equifax Data Breach” (Electronic Privacy Information Center). Accessed August 6, 2020. <https://epic.org/privacy/data-breach/equifax/>.
- “Equifax Data Breach Settlement” (Federal Trade Commission), July 11, 2019. <https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement>.
- “Equifax Data Breach Settlement | Am I Affected?” Accessed August 6, 2020. <https://www.equifaxbreachsettlement.com/>.
- “Equifax Data Breach FAQ: What Happened, Who Was Affected, What Was the Impact?” (Josh, Fruhlinger, CSO Online, February 12, 2020). <https://www.csoonline.com/article/3444488/equifax-data-breach-faq-what-happened-who-was-affected-what-was-the-impact.html>.
- “Frequently asked questions about the Equifax data breach” (Massachusetts Office of the Attorney General); <https://www.mass.gov/service-details/frequently-asked-questions-about-the-equifax-data-breach>.

Wednesday Mar 22 at 11:59pm: Deadline: paper #2 (detecting bias)

Week 8: Data Monetizing

Thu Mar 21

Interactive Lecture: economic ethics

Required reading:

- Tavani ch. 7: “Cybercrime and Cyber-Related Crimes”

Tue Mar 26

Case Study: 23 and Me monetizing genomic data

Required reading: at least two of the following:

- “Harvard genetics pioneer wants to monetize DNA with digital currency, and defeat 23andMe” <https://www.cnn.com/2018/02/08/harvard-genetics-pioneer-will-monetize-dna-with-digital-currency.html>
- “How 23andMe Is Monetizing Your DNA” <https://www.fastcompany.com/3040356/what-23andme-is-doing-with-all-that-dna>
- “23andMe Is Sharing Genetic Data with Drug Giant” <https://www.scientificamerican.com/article/23andme-is-sharing-genetic-data-with-drug-giant/>

Week 9: Cognitive Error

Thu Mar 28

Interactive Lecture: tendencies to cognitive error and virtue cultivation

Required reading:

- “Cognitive Bias” on Wikipedia; https://en.wikipedia.org/wiki/Cognitive_bias; pay particular attention to the Cognitive Bias Codex, which is an image file accessible here:

(https://en.wikipedia.org/wiki/Cognitive_bias#/media/File:Cognitive_bias_codex_en.svg).

Tue Apr 2

Case Study: Chinese social-credit surveillance

Required reading: at least two of the following:

- “Legal Documents Related to the Social Credit System”
<https://www.chinalawtranslate.com/en/social-credit-documents/>
- Government mouthpieces: the "official" far-right media is portraying it in terms of restoring morality: <https://www.globaltimes.cn/page/201905/1149741.shtml>; here is a point-by-point defensive article from China Daily: <https://global.chinadaily.com.cn/a/201911/29/WS5de0be2ea310cf3e3557af1a.html>
- SCMP is usually a good source on China, though it is now owned by Alibaba and subject to the new restrictions in Hong Kong. They have a good collection: <https://www.scmp.com/topics/chinas-social-credit-system>
- Here is a Singaporean piece from the Straits Times: <https://www.straitstimes.com/asia/east-asia/confusion-clouds-chinas-social-credit-system>
- “The Entire System Is Designed to Suppress Us’: What the Chinese Surveillance State Means for the Rest of the World” <https://time.com/5735411/china-surveillance-privacy-issues/>
- “How the West Got China's Social Credit System Wrong”
<https://www.wired.com/story/china-social-credit-score-system/>
- “The complicated truth about China's social credit system”
<https://www.wired.co.uk/article/china-social-credit-system-explained>
- “‘We are basically living naked’: The complicated truth about China's Social Credit System” <https://www.abc.net.au/news/2020-01-02/china-social-credit-system-operational-by-2020/11764740>
- “Hundreds of Chinese citizens told me what they thought about the controversial social credit system” <http://theconversation.com/hundreds-of-chinese-citizens-told-me-what-they-thought-about-the-controversial-social-credit-system-127467>
- “The complicated truth about China's social credit system”
<https://www.wired.co.uk/article/china-social-credit-system-explained>
- “China's Surveillance State Should Scare Everyone”
<https://www.theatlantic.com/international/archive/2018/02/china-surveillance/552203/>
- “China’s Social Credit System Is Actually Quite Boring”
<https://foreignpolicy.com/2021/09/15/china-social-credit-system-authoritarian/>

Part 4: Using AI to Mislead or Manipulate

Week 10: Generative AI

Thu Apr 4

Interactive Lecture: ethics of generative AI

Required reading:

- Tavani ch. 8: “Intellectual Property Disputes in Cyberspace”

Tue Apr 9

Case Study: Generative AI Survey

Required reading:

- Bring your favorite Gen AI tools to class.
- Check out <https://theresanaiforthat.com>.
- Check out <https://www.kaggle.com>.

Week 11: Mis/Disinformation

Thu Apr 11

Interactive Lecture: ethics and mis/disinformation

Required reading:

- Tavani ch. 9: “Regulating Commerce and Speech in Cyberspace”

Tue Apr 16

Case Study: Cambridge Analytica political message targeting

Required reading: for the Cambridge Analytica case study, at least two of the following:

- “The Cambridge Analytica Story, Explained” <https://www.wired.com/amp-stories/cambridge-analytica-explainer/>
- “Cambridge Analytica Whistleblower: Facebook audit was 'quite tame'” <https://money.yahoo.com/cambridge-analytica-whistleblower-facebook-company-181455061.html>
- “The scary world of who owns our data” <https://arab.news/65znq>
- “Cambridge Analytica Uncovered: Secret filming reveals election tricks” <https://www.youtube.com/watch?v=mpbeOCKZFfQ>
- “Facebook sued by Australian information watchdog over Cambridge Analytica-linked data breach” <http://www.theguardian.com/technology/2020/mar/09/facebook-cambridge-analytica-sued-australian-information-watchdog-300000-privacy-breaches>
- “Mueller questions Cambridge Analytica director Brittany Kaiser” <http://www.theguardian.com/uk-news/2019/feb/17/brittany-kaiser-trump-russia-robert-mueller-cambridge-analytica>
- “Computer science faces an ethics crisis. The Cambridge Analytica scandal proves it.” <https://www.bostonglobe.com/ideas/2018/03/22/computer-science-faces-ethics-crisis-the-cambridge-analytica-scandal-proves/lzaXxl2BsYBtwM4nxezgcP/story.html>
- “Former Cambridge Analytica exec says she wants lies to stop” <https://www.theguardian.com/uk-news/2018/mar/23/former-cambridge-analytica-executive-brittany-kaiser-wants-to-stop-lies>

- “Browser privacy: Change these settings now, whether you use Chrome, Safari or Firefox” <https://www.cnet.com/how-to/browser-privacy-change-these-settings-now-whether-you-use-chrome-safari-or-firefox/>
- “The Cambridge Analytica saga is a scandal of Facebook’s own making | John Harris” <https://www.theguardian.com/commentisfree/2018/mar/21/cambridge-analytica-facebook-data-users-profit>
- “Cambridge Analytica and Facebook: The Scandal and the Fallout So Far” <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>
- “The Facebook and Cambridge Analytica scandal, explained with a simple diagram” <https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram>
- “Revealed: the ties that bind Vote Leave's data firm to controversial Cambridge Analytica” <https://www.theguardian.com/uk-news/2018/mar/24/aggregateiq-data-firm-link-raises-leave-group-questions>
- “Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach” <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- “Our Cambridge Analytica scoop shocked the world. But the whole truth remains elusive” <https://www.theguardian.com/uk-news/2018/dec/23/cambridge-analytica-facebook-scoop-carole-cadwalladr-shocked-world-truth-still-elusive>
- “‘I made Steve Bannon’s psychological warfare tool’: meet the data war whistleblower” <https://www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-facebook-nix-bannon-trump>
- “We need tougher action against disinformation and propaganda” <https://www.brookings.edu/blog/techtank/2020/07/15/we-need-tougher-action-against-disinformation-and-propaganda/>
- “Facebook’s week of shame: the Cambridge Analytica fallout” <https://www.theguardian.com/technology/2018/mar/24/facebook-week-of-shame-data-breach-observer-revelations-zuckerberg-silence>
- “Investigators complete seven-hour Cambridge Analytica HQ search” <https://www.theguardian.com/news/2018/mar/23/judge-grants-search-warrant-for-cambridge-analyticas-offices>

Week 12: Politics

Thu Apr 18

Interactive Lecture: ethics and politics

Tue Apr 23

Case Study: Facebook used by Myanmar military to incite violence

- “‘Overreacting to failure’: Facebook's new Myanmar strategy baffles local activists” <http://www.theguardian.com/technology/2019/feb/07/facebook-myanmar-genocide-violence-hate-speech>

- “Facebook Admits It Was Used to Incite Violence in Myanmar”
<https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html>
- “A Genocide Incited on Facebook, With Posts From Myanmar’s Military”
<https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html>
- “Facebook: We didn't do enough to prevent Myanmar violence”
<https://www.cnn.com/2018/11/06/tech/facebook-myanmar-report/index.html>
- “Why Facebook is losing the war on hate speech in Myanmar”
<https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/>

Case Study: Brazilian political interference via Whatsapp

- “Disinformation Spreads on WhatsApp Ahead of Brazilian Election - The New York Times” <https://www.nytimes.com/2018/10/19/technology/whatsapp-brazil-presidential-election.html>
- “Facebook's WhatsApp flooded with fake news in Brazil election”
<https://www.reuters.com/article/us-brazil-election-whatsapp-explainer-idUSKCN1MU0UP>
- “WhatsApp fake news during Brazil election ‘favoured Bolsonaro’”
<http://www.theguardian.com/world/2019/oct/30/whatsapp-fake-news-brazil-election-favoured-jair-bolsonaro-analysis-suggests>
- “WhatsApp 'weaponised' in Brazil election” <https://www.bbc.com/news/technology-45956557>

Wed Apr 24 at 11:59pm: **Deadline for final project (policy paper)**

Week 13: Conclusion

Thu Apr 25

Case Study: digital ethics in relation to personal and family life

Tue Apr 30

Case Study: digital ethics and your professional life (with a special guest)

May 2-3: Study Period

May 6-10: Exam Week: **Final Exam (bring computers)**

Policies

Late Work

No late work will be accepted. You must submit all written assignments by the beginning of the main class meeting on the assigned date unless otherwise specified on the syllabus. Any work submitted after that time will **not** be considered. If you have outstanding circumstances that prevent you from completing the work by the assigned date, please consult with the professor.

Incompletes

Incompletes will not be given, as a matter of fairness to those who keep themselves properly organized. But health disasters do sometimes strike so reach out to the professor as soon as possible if you are in deep trouble. Remember that STH has paperwork.

Academic Code of Conduct

The BU Academic Code of Conduct may be linked to through the STH website at: www.bu.edu/sth/academic/academic-conduct. All students are required to familiarize themselves with this code, its definitions of misconduct, and its sanctions. Students should especially familiarize themselves with the section on plagiarism. On that topic:

Plagiarism

All written work in this course must be original to you. If you consult outside texts, please cite these sources in the proper format. This pertains to all external sources (books, journals, lectures, sermons, web-sites). We are required to report all suspected cases of plagiarism to the Academic Dean for review. For detailed description of the policy and procedures, please consult the STH web-site as noted above.

In some cases, students do not intend to plagiarize. Examples of plagiarism include:

failure to acknowledge the source(s) of even a few phrases, sentences, or paragraphs; failure to acknowledge a quotation or paraphrase of paragraph-length sections of a paper; failure to acknowledge the source(s) of a major idea or the source(s) for an ordering principle central to the paper's or project's structure; failure to acknowledge the source (quoted, paraphrased, or summarized) of major sections or passages in the paper or project; the unacknowledged use of several major ideas or extensive reliance on another person's data, evidence, or critical method; submitting as one's own work, work borrowed, stolen, or purchased from someone else. (Excerpted from the University of Albany's [web site](#).)

Here's a real example: in this syllabus, I use some of the language and design ideas from my colleagues at Boston University, including Prof. Kirk Wegter-McNelly and Prof. Shelly Rambo. I hereby acknowledge those professors and refer to their prior syllabi (which may not be publicly available).

Generative AI Assistance Policy

Intent:

Students should learn how to use AI text generators and other AI-based assistive resources (collectively, AI tools) to enhance rather than damage their developing abilities as writers, coders, communicators, and thinkers. Instructors should ensure fair grading for both those who do and do not use AI tools. The GAIA policy stresses transparency, fairness, and honoring relevant stakeholders such as students eager to learn and build careers, families who send students to the university, professors who are charged with teaching vital skills, the university that has a responsibility to attest to student competency with diplomas, future employers who invest in student because of their abilities and character, and

colleagues who lack privileged access to valuable resources. To that end, the GAIA policy adopts a few commonsense limitations on an otherwise embracing approach to AI tools.

Students shall:

1. Give credit to AI tools whenever used, even if only to generate ideas rather than usable text or illustrations.
2. When using AI tools on assignments, add an appendix showing (a) the entire exchange, highlighting the most relevant sections; (b) a description of precisely which AI tools were used (e.g. ChatGPT private subscription version or DALL-E free version), (c) an explanation of how the AI tools were used (e.g. to generate ideas, turns of phrase, elements of text, long stretches of text, lines of argument, pieces of evidence, maps of conceptual territory, illustrations of key concepts, etc.); (c) an account of why AI tools were used (e.g. to save time, to surmount writer's block, to stimulate thinking, to handle mounting stress, to clarify prose, to translate text, to experiment for fun, etc.).
3. Not use AI tools during in-class examinations, or assignments, unless explicitly permitted and instructed.
4. Employ AI detection tools and originality checks prior to submission, ensuring that their submitted work is not mistakenly flagged.
5. Use AI tools wisely and intelligently, aiming to deepen understanding of subject matter and to support learning.

Instructors shall:

1. Seek to understand how AI tools work, including their strengths and weaknesses, to optimize their value for student learning.
2. Treat work by students who declare no use of AI tools as the baseline for grading.
3. Use a lower baseline for students who declare use of AI tools, depending on how extensive the usage, while rewarding creativity, critical nuance, and the correction of inaccuracies or superficial interpretations in response to suggestions made by AI tools.
4. Employ AI detection tools to evaluate the degree to which AI tools have likely been employed.
5. Impose a significant penalty for low-energy or unreflective reuse of material generated by AI tools and assigning zero points for merely reproducing the output from AI tools.

This policy recognizes that:

1. This policy depends on goodwill, a sense of fairness, and honorable character.
2. Some instructors may prefer stronger restrictions on the use of AI tools and they are free to impose them so long as care is taken to maintain transparency and fairness in grading.

3. This policy takes account of the existence of subscription versions of AI tools, which are not affordable for some students; the policy may need to be revised as the differences between subscription and free versions become better understood.
4. This policy may be revised in light of other policies and novel technological developments in AI tools.

Language

This course seeks to be inclusive of people of all genders, races, cultures, abilities, and sexual orientations. Please be mindful that, when writing papers and referring to people in general, you should use terms like people, world, us, human being, humanity, etc. instead of the terms man, mankind, and men. We can talk further about what is gained and lost by this policy. If you have questions about the policy, please feel free to consult with the professor.

Disabilities

Boston University strives to be accessible, inclusive, and diverse in our facilities, programming, and academic offerings. Your experience in this class is important to the teaching staff. If you have a disability (including but not limited to learning or attention, mental health, concussion, vision, mobility, hearing, physical, or other health-related), require communication access services for the deaf or hard of hearing, or believe that you require a reasonable accommodation for another reason please meet with BU Disability and Access Services as soon as possible at the beginning of the semester to initiate disability verification and discuss accommodations that may be necessary to ensure your successful completion of course requirements. That office is at 25 Buick Street, Suite 300, and can be contacted at 617-353-3658. Requests for accommodations are then sent by that office to the Academic Dean who approves and returns them. Disability and Access Services then forwards them to the instructor.

Recordings

In this course, your image and/or voice may be recorded by the professor and/or other students. **In the case of students:** The use of technologies for photographs, and audio and video recording of lectures requires the permission of the instructor and, as applicable in relation to other classroom activities, other participants and students. Sharing these recordings is not permitted without obtaining permission from the professor and other persons being recorded. Students who have been approved for recording of lectures as a disabilities accommodation must limit their use of these recordings to personal use and are not permitted to share these recordings without permission of the professor, students, and other participants in the class. The professor always retains the right to prohibit recordings of any or all sections of lectures and will instruct students when such material/conversations are not to be recorded. All recorded material must be destroyed/erased at the end of the semester as this material contains the intellectual property of the instructor. Violation of this policy will constitute academic misconduct and be acted on according to the STH Academic Code of Conduct. **In the case of professors:** Permission must be given by students. In some cases, recordings may be used for scholarly research, presentations, publications, or future teaching. In those cases, the purpose, intended use, and scope of the recording must be disclosed. Otherwise, recordings

will only be shared with students enrolled in the course for instructional purposes and all recorded material will be deleted/erased at the end of the semester.

Grading Scale

This course will utilize the following uniform grading scale.

- $A \geq 93$
- $A- = 90-92.99$
- $B+ = 87-89.99$
- $B = 83-86.99$
- $B- = 80-82.99$
- $C+ = 77-79.99$
- $C = 73-76.99$
- $C- = 70-72.99$
- $D+ = 67-69.99$
- $D = 63-66.99$
- $D- = 60-62.99$
- $F \leq 59.99$