**DATA 521: Responsible Data Science**

**Course Leads:** Assefaw Gebremedhin and Daryl Deford

**Emails:** [assefaw.gebremedhin@wsu.edu](mailto:assefaw.gebremedhin@wsu.edu) and [daryl.deford@wsu.edu](mailto:daryl.deford@wsu.edu)

**Office:** EME B43 (Gebremedhin) and Neill 328 (Deford)

**Pre-requisites:** No pre-requisites, meant for graduate students in Data Analytics/Science and other quantitative and computational fields.

**Credit Hours: 3**

**Semester: Spring 2022**

**Classroom: Spark 233**

**Meeting times: MWF 1:10--2pm**

**Office Hours: W 2:30--3:30pm**

**Learning Management System: Canvas**

**Course Description**

This course is intended to provide graduate students who work with data an introduction to the ways that modern quantitative analysis intersects with ethical considerations. The material naturally splits into two parts, the first focusing on frameworks for analyzing ethical questions in data science and the second looking at specific topics in fairness in the context of AI and machine learning. For each of these topics, we will discuss both the broader ethical perspective using viewpoints from professional ethics as well as the technical tools and methods that have been created to address these concerns. A large emphasis in this course is also placed on preparing students to critically examine ethical dimensions of their own (future) work through relevant exercises and case studies. Students will learn to systematically apply and justify ethical frameworks to their decision-making processes.

**Course Materials**

For the first half of the course we will use the textbook Introduction to Data Ethics as a source of background material and case studies (<https://www.scu.edu/media/ethics-center/technology-ethics/IntroToDataEthics.pdf>), and we will also draw heavily on the material presented in Boyd and Crawford, “Critical Questions for Big Data” and O’Neil, “Weapons of Math Destruction.” As the second half of the course is focused on distinct subtopics in AI ethics, we will turn to a small set of relevant papers in the recent literature each week, as described in the detailed course outline later in the document. The texts are not required but recommended. Students will **not** be required to purchase a textbooks (the linked textbook they are referencing is open access and we will provide scans or links within Canvas to the other necessary readings).

Other useful resources include:

The case studies from Princeton:<https://aiethics.princeton.edu/case-studies/>

The materials prepared at Cornell: <https://docs.google.com/document/d/1GV97qqvjQNvyM2I01vuRaAwHe9pQAZ9pbP7KkKveg1o/edit>

And these materials from the University of San Francisco:<https://ethics.fast.ai/>

**Assessments**

There will be three main types of assessments in this course: discussion reflections, written assignments, and a final group analysis.

* **Discussion Reflections:** Each week, you will be required to post a reflection to Canvas summarizing and reacting to the in-class discussions and responding to the framing question that will be posed at the beginning of each class meeting.
* **Written Assignments:** Approximately other Monday an assignment will be posted to Canvas requiring you to analyze and respond to relevant case studies and to implement or apply versions of the technical content discussed in the current module. These will be due the following Wednesday (10 days after they are posted).
* **Final Project (Group Analysis):** Starting in Week 12 you will work in groups of 3 or 4 to select one of the case studies we have encountered in the semester (or with the instructor’s permission a different data project) for deeper analysis. The goal is to provide a critical assessment of the project, focusing on ethical and fairness implications. Drawing on the readings and materials from this course, you should describe the potential (or realized) ethical issues and how they might be mitigated or addressed. Your work will be summarized in a 10-page final paper and each group will give a 10-minute presentation during the final exam period, describing their work and conclusions. *The final project takes the place of the final examination for this class.*

|  |  |
| --- | --- |
| **Assessment** | **Percentage** |
| Reflections | 25% |
| Written Assignments | 50% |
| Group Analysis | 25% |

**Grading:** Letter grades will be given according to the following ranges:

A (93–100%), 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 (60–66.99%), F (less than 60%).

**Participation and attendance**

Due to the pace of the course and the range of topics that will be covered in the semester, daily attendance will be essential for your success. You are expected to have completed the required readings before the corresponding class periods. Although it is not officially a part of the course grade, missing class is likely to adversely affect your grade by impacting your understanding of the material. Class meetings will frequently incorporate activities and discussions that extend the material beyond the presentation in the readings. In particular, taking good notes of classroom discussions will be very valuable for you. It can be tempting to view recorded lectures or summaries as a substitute for personally taking notes, but a significant body of research suggests that taking notes by hand improves learning outcomes for students. Attendance will not be taken.

**Policy on Late Assignments**

Assignments or discussion reflections that are submitted after the beginning of class on Wednesday will receive one-half of the total graded points. Submissions that are more than a week late will not receive a grade. A late submission for the final group analysis will lose 10% off the final grade for each day late.

**Learning Outcomes and Assessment**

Students who successfully complete the course will be able to:

* Recognize the ethical implications of data projects
  + Assessed in Homework #1, reflections for weeks 1,3, and 6-15, and final group analysis
* Understand the role and perspective of professional ethics and ethical frameworks and how those interact with academia and industry
  + Assessed in Homework #3 and reflections for weeks 4 and 5
* Systematically evaluate potential ethical components of data research and justify appropriate responses
  + Assessed in Homework #2 and reflections for weeks 2, 3, and 6-15, and final group analysis
* Situate their work in relevant legal and cultural frameworks
  + Assessed in Homework #2 and #3, reflections for weeks 2-5, and final group analysis
* Promote privacy and security in data acquisition and storage practices
  + Assessed in Homework #2 and #3, reflections for weeks 2-5, and final group analysis
* Understand and evaluate modern work relating technical concepts and ethical concerns in:
  + Feedback loops
    - Assessed in Homework 4 and week 6 reflections
  + Ranking systems
    - Assessed in Homework 4 and week 7 reflections
  + AI Fairness
    - Assessed in Homework 5 and week 8 and 9 reflections
  + Model Bias and Interpretability
    - Assessed in Homework 6 and week 10 and 12 reflections
  + Differential Privacy
    - Assessed in Homework 2 and week 3 reflections
  + Factor Analysis
    - Assessed in Homework 6 and week 11 reflections
  + Social Media and Information Diffusion
    - Assessed in Homework 7 and weeks 12-15 reflections
  + Depending on choice of topic, students will examine and be assessed on some but not all of these concepts in their final group analysis

**Collaboration Policy**

For the assignments you are encouraged (and sometimes required) to work with other students in the class. However, the work that you submit should be your own, and in particular, it should be written in your own words and communicated with your own understanding of the solution. If you do collaborate, please list the names of the other students you worked with on your submission. You may be asked to explain your reasoning individually to obtain full credit. Obtaining solutions from external sources like chegg or coursehero for course problems will be considered a violation of the academic integrity policy with consequences described below.

**Weekly Topic Outline**

Note: Below is a detailed outline of the topics in each week, accompanied by representative readings for that week.

**Part I (Broad ethical and legal considerations in data science)**

**Week 1 Topic:** *Intro to ethical frameworks and professional ethics*

**Week 1 Readings:** Examples of professional guidelines in the field:

* <http://www.datascienceassn.org/sites/default/files/datasciencecodeofprofessionalconduct.pdf> (*Code of Professional Conduct*)
* <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication__1_.pdf> (*Data Science Ethical Framework*}
* <https://www.amstat.org/asa/files/pdfs/EthicalGuidelines.pdf> (*Ethical Guidelines for Statistical Practice*)
* <http://www.datascienceassn.org/code-of-conduct.html> (*Data Science Association)*
* <https://www.uio.no/studier/emner/sv/oekonomi/ECON4135/h09/undervisningsmateriale/FinancialModelersManifesto.pdf> (*Modeler’s Hippocratic Oath*)
* <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication__1_.pdf> (*UK Data Science Ethical Framework*)
* <https://aoir.org/reports/ethics3.pdf> (Ethical Decision-Making and Internet Research (AoIR)
* <https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx> (Ethical Guidelines for Statistical Practice (ASA)
* Data Science Ethos Lifecycle (ADSA)

**Week 2 Topics**:

* *Legal frameworks (and interventions)*
  + Official use of data (surveillance)
  + Comparison of the GDPR (European laws) to the ethical frameworks in terms of both rights of individuals and responsibilities of researchers
  + State-by-state rulemaking (use of surveillance cameras by police, 2020 Illinois banned AI analyzed video interviews)
  + Constraints on Human Subject Research (IRBs) or confidential data as mediated by ORSO at WSU
  + Official rules about data
  + Intersections with (applied) ethics
* *Intellectual Property*
  + Copyright (Is data copyrightable? What outputs of data analytics work are copyrightable?)
  + Licensing (How do data licenses differ from software licenses - modern open data practices)
  + Academic products

**Week 2 Readings:**

* Lippert-Rasmussen, “‘We Are All Different’: Statistical Discrimination and the Right to Be Treated as an Individuals”
* Jones, “The right to a human in the loop

**Week 3 Topics:**

* *Data Privacy* (applying techniques in previous two weeks to analyze specific examples)
  + Dataset release and breaches case studies. Examples: OkCupid data release; Enron emails; Kickstarter data case study; T-Mobile 2021 and Target 2015.
* *Differential Privacy*
  + Basic mathematical definitions (\varepsilon-DP Cynthia Dwork, TopDown algorithm Moon Duchin)
  + Inputs and outputs of these algorithms
  + Implicit notion of safety or privacy (how does that interact with ethical frameworks)
  + Apply the method to a dataset
  + 2021 census applied differential privacy to block level data - analysis of the impacts (costs vs. benefits)

**Week 3 Readings:**

* Dwork, Hardt, Pitassi, Reingold, and Zemel, “Fairness Through Awareness”
* <https://imai.fas.harvard.edu/research/files/Harvard-DAS-Evaluation.pdf>
* <https://drive.google.com/file/d/1ov8evbT5bSCiU33UzssFPEkIPfa5uVFw/view>
* <https://mggg.org/d>p

**Week 4 Topic:** *Academic Ethics* (with specific application to data perspectives)

* Publication ethics
* Authorship practices in different fields
* Conferences vs. journals
* Cultural aspects of data in academia (context of living between different fields)
* Data storage and retention
* IRBs
* Openness and data publication

**Week 4 Readings:**

* On Being a Scientist (National Academy of Sciences)
* http://math.wsu.edu/faculty/ddeford/ethics.php

**Week 5 Topic:** *Financial industry (consulting) ethics*

* Broad ethical considerations
* Day-to-day roles
* Impacts of models on large scale social institutions (2008 crash)
* Data/modeling citizenship - what should/do analysts contribute to society?

**Week 5 Readings:**

* <https://www.uio.no/studier/emner/sv/oekonomi/ECON4135/h09/undervisningsmateriale/FinancialModelersManifesto.pdf> (*Modeler’s Hippocratic Oath*)
* Ethics and Quantitative Finance (West)

**Part II (Ethics and fairness in AI/ML):**

**Week 6 Topic:** *Feedback loops*

* Recommender systems
* Ride sharing apps (5-star ratings)
* Impact Factors (preferential attachment)
* Post-implementation evaluation (e.g. in public health settings)

**Week 6 Readings:**

* Intro and Chapter 1 Weapons of Math Destruction (O’neil)

**Week 7 Topic:** *Scoring systems* (ranking individuals and entities)

* Credit scores
* Teacher evaluation metrics
* COMPAS recidivism algorithm
* University rankings (US News, CS Rankings, etc.)

**Week 7 Readings:**

* Kleinberg, Mullainathan, and Raghavan, “Inherent Trade-Offs in the Fair Determination of Risk Scores”
* Northpointe, *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*
* Hurley and Adebayo, “Credit Scoring in the Era of Big Data”

**Weeks 8 and 9 Topic:** *Measuring fairness in machine learning and AI*

* Definitions of neutrality
* Bias Metrics
* Kleinberg’s work to evaluate how methods are doing along these axes
* Algorithmic auditing
* Implicit bias analysis (blind orchestra auditions)
* Racial biases in facial recognition

**Weeks 8 and 9 Readings:**

* Friedman and Nissenbaum, “Bias in Computer Systems” Sandvig, Hamilton, Karahalios, and Langbort, “Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms”
* Datta, Tschantz, and Datta, “Automated Experiments on Ad Privacy Settings”

**Week 10 Topic:** *Impacts of training data*

* Problems caused by datasets lacking diversity
* How to address this in algorithm/experiment design
* Representational Harms
* Adversarial learning examples (GANs)
* Case studies

**Week 10 Readings**

* Buolamwini, “Algorithms Aren’t Racist. Your Skin Is just too Dark”
* Hassein, “Against Black Inclusion in Facial Recognition”
* Agüera y Arcas, Mitchell, and Todorov, “Physiognomy’s New Clothes”

**Week 11 Topic:** *Factor Analysis* (Correlation vs. Causation)

* Historical context
* Race and IQ
* STS readings (Andrew Hacker)
* Causality models
* Bayesian (Graphical) models

**Week 11 Readings:**

* Stephen Jay Gould, [Factor Analysis and the Reification of Intelligence](https://sites.tufts.edu/models/files/2019/01/gould-mismeasure.pdf)

**Weeks 12 and 13 Topic:** *Transparency and interpretability*

* Interpretable AI
* Deep learning and interpretability
* How to design systems that communicate effectively to stakeholders
* Clarity about uses of raw data

**Weeks 12 and 13 Readings:**

* Lipton, “The Mythos of Model Interpretability”
* Doshi-Velez and Kim, “Towards a Rigorous Science of Interpretable Machine Learning”

**Week 14 Topic:** *Social media (filter bubbles)*

* Facebook/Cornell studies 2013
* Network propagation (diffusion of opinions)
* How effectively messages can spread

**Week 14 Readings:**

* Pariser, “Beware Online ‘Filter Bubbles’”
* Gillespie, “The Relevance of Algorithms”

**Week 15 Topic:**  *Disinformation/Misinformation*

* Intentional spreading vs. structural effects
* AI impacts
* Botnets and adversarial actors
* Youtube recommendations and radicalization

**Week 15 Readings:**

* Mark Ledwich, Anna Zaitsev, “Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization”

**COVID-19 Policy**

Students are expected to abide by all current COVID-19 related university policies and public health directives. These directives may be adjusted to respond to the evolving COVID-19 pandemic. Directives may include, but are not limited to, compliance with WSU’s COVID-19 vaccination policy, wearing a cloth face covering, physically distancing, and sanitizing common use spaces. All current COVID-19 related university policies and public health directives are located at https: //wsu.edu/covid-19/. Students who do not comply with these directives may be required to leave the classroom; in egregious or repetitive cases, student non-compliance may be referred to the Center for Community Standards for action under the Standards of Conduct for Students.

**Academic Integrity**

Academic integrity is the cornerstone of higher education. As such, all members of the university community share responsibility for maintaining and promoting the principles of integrity in all activities, including academic integrity and honest scholarship. Academic integrity will be strongly enforced in this course. Students who violate WSU’s Academic Integrity Policy (identified in Washington Administrative Code (WAC) 504-26-010(4) will fail the assignment implicated, will not have the option to withdraw from the course pending an appeal, and will be reported to the Center for Community Standards.

Cheating includes, but is not limited to, plagiarism and unauthorized collaboration as defined in the Standards of Conduct for Students, WAC 504-26-010(3). Read and understand all of the definitions of cheating given here <http://app.leg.wa.gov/WAC/default.aspx?cite=504-26-010>. If you have any questions about what is and is not allowed, ask your course instructor.

If you wish to appeal instructor’s decision relating to academic integrity, please use the form available at <http://communitystandards.wsu.edu>. Make sure you submit your appeal within 21 calendar days of the instructor’s decision.

**Students with Disabilities**

Reasonable accommodations are available for students with documented disabilities or chronic medical conditions. If you have a disability and need accommodations to fully participate in this class, please visit your campus Access Center website (websites listed below) to follow published procedures to request accommodations. Students may also call or email the Access Center to schedule an appointment with an Access Advisor. All disability related accommodations are to be approved through the Access Center. It is a university expectation that students with approved accommodations visit with instructors (in person or via Zoom) within two weeks of requesting their accommodations to discuss logistics.

For more information contact a Disability Specialist on your home campus:

* Pullman or WSU Online: 509-335-3417 <http://accesscenter.wsu.edu>, Access.Center@wsu.edu
* Spokane: <https://spokane.wsu.edu/studentaffairs/access-resources/>
* Tri-Cities: <http://www.tricity.wsu.edu/disability/>
* Vancouver: 360-546-9138; <https://studentaffairs.vancouver.wsu.edu/student-wellness-center/access-center>

**Accommodation for Religious Observances or Activities**

Washington State University reasonably accommodates absences allowing for students to take holidays for reasons of faith or conscience or organized activities conducted under the auspices of a religious denomination, church, or religious organization. Reasonable accommodation requires the student to coordinate with the instructor on scheduling examinations or other activities nec- essary for course completion. Students requesting accommodation must provide written notifi- cation within the first two weeks of the beginning of the course and include specific dates for absences. Approved accommodations for absences will not adversely impact student grades. Ab- sence from classes or examinations for religious reasons does not relieve students from responsi- bility for any part of the course work required during the period of absence. Students who feel they have been treated unfairly in terms of this accommodation may refer to Academic Regu- lation 104 – Academic Complaint Procedures. See also Academic Regulation 82, available at <https://registrar.wsu.edu/academic-regulations/>

**Safety and Emergency Notification**

Classroom and campus safety are of paramount importance at Washington State University, and are the shared responsibility of the entire campus population. WSU urges students to follow the “Alert, Assess, Act” protocol for all types of emergencies and “Run, Hide, Fight” response for an active shooter incident. Remain ALERT (through direct observation or emergency notification), ASSESS your specific situation, and act in most appropriate way to assure your own safety (and the safety of others if you are able).

Please sign up for emergency alerts on your account at MyWSU. For more information on this subject, campus safety and related topics, please view the FBI’s Run, Hide, Fight video (https://www. fbi.gov/about-us/cirg/active-shooter-and-mass-casualty-incidents/run-hide-fight-video) and visit the WSU safety portal (<https://faculty.wsu.edu/classroom-safety>).

Full details can be found at <https://provost.wsu.edu/classroom-safety/>

**Academic Dates and Deadlines**

Students are encouraged to refer to the academic calendar often to be aware of critical deadlines throughout the semester. The academic calendar can be found at <http://registrar.wsu.edu/academic-calendar>

###### **Expectations for Student Effort**

For each hour of lecture equivalent, students should expect to have a minimum of two hours of work outside class.