STOR 390 – Fall 2024 Machine Learning: Ethics and Society

Course Overview

Modern statistics introduces a number of normative issues. Concerns surrounding algorithmic bias, data privacy, and methodological transparency arise from our methods, but our methods are not capable of *fully* ameliorating them. A more complete answer must be sought in moral philosophy. This *statistics course* is positioned at the intersection these disciplines, and is as concerned with *why* to implement a method as how. It will combine rigorous analysis of modern data sets alongside open-ended philosophical discussion. Ultimately, students leaving this class will be able to defend when their methods are not only technically sound but morally just.

Learning Objectives

This course will enable you to:

- explore ethical dilemmas arising from modern statistical methods;
- consider technical solutions to such dilemmas;
- leverage philosophical principals to adjudicate instances where technical solutions remain ambiguous; and
- communicate and present integrated results and position.

Course Topics

Statistical: Classification, Sensitivity Analysis, Federated Learning, Causal Inference, Open-Source Coding

Philosophical: Deontology, Consequentialism, Virtue Ethics, Harm Principle, Informed Consent Basic Information

Instructor	Andrew Ackerman (Andy)
	Email: <u>atacker@email.unc.edu</u>
	Office: Hanes B-7
	Office Hours: TTh 9:30-11:00 am, or by appointment
Contact	When emailing me, please simply call me Andy, and include "390" in the subject line. You may
	include screenshots or photos of your scratch work.
Course	In-Person unless otherwise advised. In the case of a transition to purely virtual instruction,
Mode	lectures will continue synchronously via zoom and exam policies will be made explicit.
Lecture	TTh 8:00 – 9:15 am; Hanes 130
Course	There is no set textbook for this course. In writing my lectures, I most closely follow reference
materials	16, but this is by no means a required text.
	All course materials are open-source and legally made available on the Canvas page. These
	appear in the form of R-Markdown and reference texts.
Credit hours	3
Prerequisites	STOR 120 or comparable
Target	This course is ideally suited for STAN majors and minors interested in a career that involves real
Audience	data analysis. In principle this class will reference concepts in linear modelling, classification,
	linear programming, and causal inference. As such, a perspective student should have at

minimum taken STOR 120. Additionally, classes such as STOR 320 and STOR 455 may be useful but not absolutely necessary for success in this class.

Final Grade Computation

Assessments	Date	Percent
Class Participation	Daily	15%
Homework	Biweekly	20%
Quizzes	4 in total	10%
Midterm Project	Oct 24	25%
Final Project	Dec. 3 (LDOC)	30%

А	93 or above	С	73 to 76.99
A-	90 to 92.99	C-	70 to 72.99
B+	87 to 89.99	D+	67 to 69.99
В	83 to 86.99	D	63 to 66.99
В-	80 to 82.99	D-	60 to 62.99
C+	77 to 79.99	F	Below 60

Course Assessments

Class Participation	Insofar as this class will be more Socratic and conversational than a typical "stand- and-deliver" lecture, regular class participation is paramount. This will be assessed via a combination of attendance and engagement in class discussion.
Class Readings	The listed readings below are meant to help stimulate our discussions or introduce you to a statistical concept. They are required readings, both in the sense that I will assume prior knowledge while referencing them in class, and in the sense that conceptual comprehension of these texts may be assessed on quizzes and homework.
Homework	Homework assignments will be given roughly once every other week. These are meant to exercise the technical skill set we develop in class while also asking you to engage with the philosophical principles that uphold them. Most assignments will entail a few computational/modelling questions followed by a short defense (or critique) of the analysis itself.
Projects	There will be one project split into two parts – a midterm and final. In essence, this is meant to be a careful evaluation of a contemporary scholarly paper in either statistics, machine learning, economics, or machine learning. <u>Google scholar</u> and <u>Scopus</u> are good places to start! More specifically, the midterm should identify a paper whose methods (either explicitly or tacitly) give rise to a normative question. Along with elaborating on the relevance of this normative question, this portion of the project should also include a concise summary of the method, results, and impact on the given field.

	The final should aim to defend or critique this same paper using principles and techniques discussed in class. Ideally, this should include both verbal and mathematical argumentation. As to whether the mathematical argument is a formal proof, simulation, real data analysis, etc., this is left to the discretion of the student. Both midterm and final will be assessed for their clarity, professionalism, and persuasive appeal. In conjunction with the github mini-series at the end of the course, the finalized project should be compiled and presented as a public repository. Precise details of each leg of this project, as well as sample rubrics, are included on the course site.
Quizzes	There will be four quizzes in total. These will be aimed at assessing conceptual techniques from the class as well as your ability to defend a position involved in a given dilemma.

Course Policies

Teaching Philosophy	My general approach aims to make two principles abundantly clear: there are no (sincere) bad questions, and education is only one facet of your life. To the first point, if you ask me a question out of genuine curiosity or confusion, I guarantee to respect and attend to it regardless of the level of complexity. Note, I do not claim to always have the answer, and certain questions may require an outside (office hours) conversation. But my guarantee is to provide an answer when I am able and search with you when unable. The flip side of this guarantee is your willingness to keep me accountable. Ask questions. If you are confused, please reach out for help. If in-class questions are too daunting, take advantage of office hours or electronic communication. As to the second point, while I am ultimately a very ardent advocate of the merits of education, I want to reiterate that no single facet of your life – perhaps least of which performance in STOR 390 – fully defines you as a person. This is not to say that this, or any class, is unimportant, rather that this class is meant to enable and instruct not to define.	
Honor Code	It is my aim to make my expectations for acceptable and original work as clear and explicit as possible. Gray areas frustrated me as a student, so I aim to prevent them as an instructor. To that end, students are bound by UNC's <u>honor code</u> in taking exams and in written work, and the submission of said work signifies understanding and acceptance of those requirements. More specific to this course, collaboration is not only permitted but <i>encouraged</i> on homework and homework alone. Projects and quizzes, by contrast, will be completed individually. I recognize that exams can be immensely stressful, and I do not intend to add to that burden. That said, my job is to cultivate and assess your <i>personal</i> understanding. Insofar as individual exams serve this purpose, they will be preferred. Please consult with me if you have any questions about the honor code. Asking for clarification in advance is far preferred to asking for forgiveness post-facto.	
Attendance Policy	Attendance is required in accordance with the class participation policy.	

Mask Requirements	In accordance with the UNC mask policy, masks are optional for all students and instructors.
Technology Use	 We will frequently be using R in class. As such, laptops are not only permitted but necessary. Technology can support student learning, but it can also become a distraction. You should not use laptops or other devices for entertainment during class and should not display any material on the laptop which may be distracting or offensive to your fellow students. Laptops should only be used for legitimate classroom purposes, such as coding, taking notes, or downloading class information from Sakai. E-mail, messaging, surfing the Internet, reading the news, or playing games are not considered legitimate classroom purposes. Such inappropriate laptop use is distracting to those seated around you and is unprofessional.
Syllabus changes	I reserve the right to make changes to the syllabus, including test dates and grade computation. These changes will be announced as early as possible so that students can adjust their schedules.

University Resources for Students

Learning Center	The <u>UNC Learning Center</u> is a great resource both for students who are struggling in their courses and for those who want to be proactive and develop sound study practices to prevent falling behind. They offer individual consultations, peer tutoring, academic coaching, test prep programming, study skills workshops, and peer study groups.
Accessibility Resources	UNC-Chapel Hill facilitates the implementation of reasonable accommodations for students with learning disabilities, physical disabilities, mental health struggles, chronic medical conditions, temporary disability, or pregnancy complications, all of which can impair student success. See the ARS website for contact and registration information: <u>https://ars.unc.edu/about-ars/contact-us</u>
Counseling and Psychological Services	CAPS is strongly committed to addressing the mental health needs of a diverse student body through timely access to consultation and connection to clinically appropriate services, whether for short or long-term needs. Go to their website: <u>https://caps.unc.edu/</u> or visit their facilities on the third floor of the Campus Health Services building for a walk-in evaluation to learn more.
Title IX	Any student who is impacted by discrimination, harassment, interpersonal (relationship) violence, sexual violence, sexual exploitation, or stalking is encouraged to seek resources on campus or in the community. Please contact the Director of Title IX Compliance (Adrienne Allison - <u>Adrienne.allison@unc.edu</u>), Report and Response Coordinators in the Equal Opportunity and Compliance Office (<u>reportandresponse@unc.edu</u>), Counseling and Psychological Services (confidential), or the Gender Violence Services Coordinators (<u>gvsc@unc.edu</u> ; confidential) to discuss your specific needs. Additional resources are available at safe.unc.edu.

Ombuds Office	The Ombuds Office (https://ombuds.unc.edu/) is an impartial, confidential, and independent office within UNC that offers a safe space to any UNC members to discuss sensitive issues (for example, issues relating to power dynamic). They do not take sides and maintain complete confidentiality. All conversations with the Ombuds Office are "off the record."		
Andy's Teaching Mentors	As a graduate Teaching Fellow, I am mentored by STOR professors Dr. Mario Giacomazzo (<u>mgiacoma@email.unc.edu</u>) and Dr. Jeff McLean (<u>mclean@unc.edu</u>). If you have an issue with the class that you feel uncomfortable raising with me, you may raise it with them confidentially. The Ombuds Office (above) is also an option.		

Rough Course Outline (updated 7/15/24)

	-				
Class	Day	Date	Topics	Due	Reading
1	Т	20-Aug	Introduction/Syllabus/ Motivating Ethical Dilemmas		2,15, 18
			Moral Philosophy Primer		
2	Th	22-Aug	Consequentialism		3 pg. 4-11, 4 pg 31-37, 48-52
3	Т	27-Aug	Deontology		4 pg 120-124
4	Th	29-Aug	Virtue Ethics		
	Т	3-Sept	No Class – Wellness Day		
6	Th	5-Sept	Pitfalls of Normative Theories	HW 1	
7	Т	10-Sept	Justifiability		17, 11 (sections 1, 3.1)
			Reproducibility		
8	Th	12-Sept	Introduction to Github (Building Repo/ReadMe)		
9	Т	17-Sept	Introduction Github (Branching/Forking/Merging)	Quiz 1	
			Algorithmic Bias		
10	Th	19-Sept	Classification (Introduction via Decision Theory)		8 (Ch. 3)
11	Т	24-Sept	Classification (KNN)		
12	Th	26-Sept	Classification (SVM)	HW 2	
13	Т	1-Oct	Classification (Decision Trees)		
14	Th	3-Oct	Classification (Logistic Regression)		
15	Т	8-Oct	Classification (Logistic Regression)	HW 3	
16	Th	10-Oct	Philosophical Measures of Fairness/Justice	Quiz 2	4 pg. 62-66, pg. 149-153

17	Т	15-Oct	Statistical Measures of Fairness		7 pg 4-10
	Th	17-Oct	No Class – Fall Break		
18	Т	22-Oct	Non-arbitrary Discrimination		8 (Ch 4 pg 76-80)
19	Th	24-Oct	Case Study – COMPAS	Midterm Project	1
			Data Privacy		
20	Т	29-Oct	Informed Consent/Autonomy/Harm Principle	HW 4	13, 14 (Sections 1,2), 10 (Section 3.6)
21	Th	31-Oct	Federated Learning		5, 6 (section 1)
22	Т	5-Nov	Federated Learning		
23	Th	7-Nov	Randomized Response Differential Privacy	HW 5	
24	Т	12-Nov	NDA/Data Purging	Quiz 3	
			Explainable AI: Transparency and the "Black Box	r"	
25	Th	14-Nov	Explainable AI	HW 6	9 pg. 1-6
26	Т	19-Nov	Explainable AI		12
Precise Interpretation					
27	Th	21-Nov	Simpson's Paradox	Quiz 4	
28	Т	26-Nov	Confounding/Batch Effects	HW 7	
	Th	28-Nov	No Class – Thanksgiving Recess		
29	Т	3-Dec	Moral Absolutes	Final Project	

Class Readings (updated 9/5/24):

- 1. Larson, J., Angwin, J., Kirchner, L., & Mattu, S. (2016, May 23). *How we analyzed the compas recidivism algorithm*. ProPublica. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
- 2. Saltz, J. S., & Dewar, N. (2019). Data science ethical considerations: a systematic literature review and proposed project framework. *Ethics and Information Technology*, *21*, 197-208.
- 3. May, T. (2021a). A decent life: Morality for the rest of us / todd may. The University of Chicago Press.
- 4. Sandel, M. J. (2010). *Justice: What's the right thing to do?* Farrar, Straus and Giroux.
- T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," in *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50-60, May 2020, doi: 10.1109/MSP.2020.2975749.
- McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273-1282). PMLR.

- 7. Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. ACM Computing Surveys (CSUR), 55(3), 1-44.
- 8. Barocas, S., Hardt, M., & Narayanan, A. (2017). Fairness in machine learning. Nips tutorial, 1, 2017
- 9. Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(5), e1424.
- Brink, David, "Mill's Moral and Political Philosophy", *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = https://plato.stanford.edu/archives/fall2022/entries/mill-moral-political/>.
- Ashford, Elizabeth and Tim Mulgan, "Contractualism", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2018/entries/contractualism/>.
- 12. Garret, Brandon Rudin Cynthia (2023). Interpretable Algorithmic Forensics. Proceedings of the National Academy of Sciences
- 13. Berreby, D. (2017, March 3). *Click to agree with what? no one reads terms of service, studies confirm*. The Guardian. https://www.theguardian.com/technology/2017/mar/03/terms-of-service-online-contracts-fine-print
- Powles J, Hodson H. Google DeepMind and healthcare in an age of algorithms. Health Technol (Berl). 2017;7(4):351-367. doi: 10.1007/s12553-017-0179-1. Epub 2017 Mar 16. PMID: 29308344; PMCID:PMC5741783.
- 15. Zimmerman, J. (2023, August 30). *Opinion | here's my Al policy for students: I don't have One*. The Washington Post. https://www.washingtonpost.com/opinions/2023/08/29/ai-student-policy-chatgpt-college/
- 16. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (1st ed.) [PDF]. Springer.
- 17. Vardeman, S., Morris, M., (2003) Statistics and Ethics, The American Statistician, 57:1, 21-26, DOI: <u>10.1198/0003130031072</u>
- 18. Skipper, R., Currie, R., Elliot, D., Feldman, A., Jones, C., McWilliams, S., Wietrzykowski, M., (2024) 28th APPE Intercollegiate Ethics Bowl National Championship, Case 2: The Last Invention