

I310D - Introduction to Human-Centered Data Science
Monday/Wednesday, 9:00AM
SZB 3.814

Instructor: Maggie Engler

Pronouns: she/her

Email: maggie.engler@utexas.edu

TA: Daniela Lizarazo

Pronouns: they/them

Email: dlizarazo@utexas.edu

Office Hours

Maggie Engler: Office hours by appointment only.

Daniela Lizarazo: TBA

Communication and Asking for Help

Please ask all questions that are applicable to the entire class on Canvas, so that others may benefit from the discussion. Only use email for questions unique to individual circumstances; in those cases, please address all questions to *both* maggie.engler@utexas.edu and dlizarazo@utexas.edu.

Course Description

I310D: Introduction to Human-Centered Data Science is a survey course that introduces students to the theory and practice of data science through a human-centered lens, with emphasis on how design choices influence algorithmic results. Students will gain comfort and facility with fundamental principles of data science including research ethics, privacy, bias, fairness, transparency, accountability, reproducibility, interpretability, and societal implications.

Prerequisites

I301: Introduction to Informatics is a prerequisite or corequisite for this course.

Learning Outcomes

Following the completion of the course, students should be able to:

- Explain the lifecycle and requirements for conducting human-centered data science
- Recognize appropriate data-driven strategies to apply in a diverse array of circumstances based on the problem statement, conditions and constraints
- Identify existing or potential ethical, technical, or logical issues within data science applications, and reflect on their societal impact

Instruction Modality

Class meetings will be in person, with some exceptions and dependent on the state of the COVID-19 pandemic. If we are unable to meet in person, classes will be held virtually via Zoom.

Accommodations for Students with Disabilities

The university is committed to creating an accessible and inclusive learning environment consistent with university policy and federal and state law. Please let me know if you experience any barriers to learning so I can work with you to ensure you have equal opportunity to participate fully in this course. If you are a student with a disability, or think you may have a disability, and need accommodations please contact Services for Students with Disabilities (SSD). Please refer to SSD's website for contact and more information: <http://diversity.utexas.edu/disability/>. If you are already registered with SSD, please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations and needs in this course.

Required Materials

There is no required textbook for this course; all assigned readings will be available online at no cost. You may find the readings in the course outline section below.

Required Devices

This course is device-agnostic, but we do require the installation of Python and Jupyter notebooks.

Major Assignments

Reading Reflections (*due weekly on Mondays*)

Assignment 1 - Data Collection and Curation (*due September 19*)

Assignment 2 - Data Bias (*due October 17*)

Final Project (*plan due October 31; presentation due November 30; report due December 4*)

Late Work and Missed Work

In an effort to accommodate any unexpected personal events, I have enacted a grace policy of two days for this course. You do not have to utilize this policy, but if you find yourself struggling with unexpected personal events, I encourage you to email me as soon as possible (in advance of the due date) to notify me that you are using our grace policy. You may either have a two-day grace period for one assignment, **or** you may have 2 one-day extensions for two different assignments.

The only absences that will be considered excused are for religious holidays or extenuating circumstances due to an emergency. If you plan to miss class due to observance of a religious holiday, please let us know at least two weeks in advance. You will not be penalized for this absence, although you will still be responsible for any work you will miss on that day if applicable. In the event of an unexcused absence, we do not guarantee the opportunity to make up missed in-class work, but one may be granted. Check with us for details or arrangements.

Grading Policies

Course grades will be made up of the following components. Final letter grades will be awarded according to the grade cutoffs below, including pluses and minuses.

Grade Component	Percentage
Class participation and attendance	20%
Reading reflections	15%
Final project	35%
Other assignments	30%

Grade Breaks

Grade	Cutoff
A	94%
A-	90%
B+	87%
B	84%
B-	80%
C+	77%
C	74%
C-	70%
D+	67%
D	64%
D-	60%
F	<60%

Course Outline

Week 1 - Introduction to Human-Centered Data Science

Reading: S. Barocas and H. Nissenbaum (2014) "[Big Data's End Run Around Consent and Anonymity.](#)" In *Privacy, Big Data and the Public Good* Eds. J. Lane, V. Stodden, S. Bender, H. Nissenbaum, Cambridge: Cambridge University Press.

Week 2 - Privacy and Consent

Readings: Beckett, Lois. [“Everything We Know About What Data Brokers Know About You”](#). *ProPublica*, June 2014.

Week 3 - Data Summary and Visualization

Readings: Kitzes, J., Turek, D. and Deniz, F (Eds.). *The Practice of Reproducible Research*. [Chapter 2: Assessing Reproducibility](#) and [Chapter 3: The Basic Reproducible Workflow](#).

Week 4 - Data Collection and Reproducibility

Readings: Gray, Mary L. and Suri, Siddharth. [“What it's really like to be one of the ghost workers on Amazon's Mechanical Turk.”](#) *Fast Company*, May 2019.

Week 5 - Crowdsourcing

Readings: Buolamwini, J. and Gebru, T. [“Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.”](#) In *Proceedings of Machine Learning Research* 81:1-15, 2018. Conference on Fairness, Accountability and Transparency.

Week 6. - Data Bias

Readings: Wang, Tricia. [“Why Big Data Needs Thick Data.”](#) *Ethnography Matters*, 2016.

Week 7 - Qualitative Research and Mixed Methods

Readings: Raschka, Sebastian. [Chapter 1: Introduction to Machine Learning and Deep Learning](#).

Week 8 - Machine Learning Algorithms

Readings: Donovan, J., Caplan, R., Matthews, J., & Hanson, L. (2018). [“Algorithmic accountability: A primer.”](#) *Data & Society*, 501(c).

Week 9 - Fairness, Accountability and Transparency

Readings: Diakopoulos, N. (2014). [“Algorithmic accountability reporting: On the investigation of black boxes.”](#) *Tow Center for Digital Journalism*, 1–33.

Week 10 - Human-Centered Algorithmic Auditing

Readings: Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). [“Local Interpretable Model-Agnostic Explanations: An Introduction.”](#) O'Reilly.

Week 11 - Model Interpretability

Readings: Russell, Stuart. [“How to Stop Superhuman AI Before It Stops Us.”](#) *New York Times*, Oct. 2019.

Week 12 - Unsupervised Methods

Readings: Mitchell, Melanie. [“We Shouldn't Be Scared by 'Superhuman AI'.”](#) *New York Times*, Oct. 2019.

Week 13 - Database Design

Readings: None

Week 14 - FALL BREAK

Readings: None

Week 15 - Careers in Human-Centered Data Science

Readings: None

Week 16 - Final Presentations

Readings: None

Week 17 - NO FINAL EXAM**Course Policies and Disclosures***Academic Integrity*

Students who violate University rules on academic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on academic dishonesty will be strictly enforced. For further information, please visit the Student Conduct and Academic Integrity website at <http://deanofstudents.utexas.edu/conduct>.

Confidentiality of Class Recordings

In the event that class should be recorded, class recordings are reserved only for students in this class for educational purposes and are protected under FERPA. The recordings should not be shared outside the class in any form. Violation of this restriction by a student could lead to Student Misconduct proceedings.

Religious Holy Days

By [UT Austin policy](#), you must notify me of your pending absence as far in advance as possible of the date of observance of a religious holy day. If you must miss a class, an examination, a work assignment, or a project in order to observe a religious holy day, you will be given an opportunity to complete the missed work within a reasonable time after the absence.

Names and Pronouns

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. I will gladly honor your request to address you by your chosen name and by the gender pronouns you use. Class rosters are provided to the instructor with the student's chosen (not legal) name, if you have provided one. If you wish to provide or update a chosen name, that can [be done easily at this page](#), and you can [add your pronouns to Canvas](#).

Basic Needs Security

Any student who faces challenges securing their food or housing and believes this may affect their performance in the course is urged to contact the Dean of Students for support. UT maintains the [UT Outpost](#) which is a free on-campus food pantry and career closet.

Mental Health Support

I urge students who are struggling for any reason and who believe that it might impact their performance in the course to reach out to me if they feel comfortable. This will allow me to provide any resources or accommodations that I can. If immediate mental health assistance is needed, call the Counseling and Mental Health Center (CMHC) at 512-471-3515 or you may also contact Bryce Moffett, LCSW (iSchool CARE counselor) at 512-232-2983. Outside CMHC business hours (8a.m.-5p.m., Monday-Friday), contact the CMHC 24/7 Crisis Line at 512-471-2255.

Land Acknowledgement

I would like to acknowledge that we are meeting on the Indigenous lands of Turtle Island, the ancestral name for what now is called North America. Moreover, I would like to acknowledge the Alabama-Coushatta, Caddo, Carrizo/Comecrudo, Coahuiltecan, Comanche, Kickapoo, Lipan Apache, Tonkawa and Ysleta Del Sur Pueblo, and all the American Indian and Indigenous Peoples and communities who have been or have become a part of these lands and territories in Texas.