I310D - Introduction to Human-Centered Data Science
Monday/Wednesday/Friday, 9:00AM
Parlin Hall (PAR) 208

Instructor: Maggie Engler
Pronouns: she/her/hers
Email: maggie.engler@utexas.edu

TA: Li Shi
Pronouns: she/her/hers
Email: lilylashi@utexas.edu

Office Hours
Maggie Engler: Office hours by appointment. Please email maggie.engler@utexas.edu to schedule.
Li Shi: Office hours by appointment. Please email lilylashi@utexas.edu to schedule.

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 lilylashi@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 February 11)*
Assignment 2 - Data Bias *(due March 11)*
Final Project *(plan due April 8; report due April 22; presentation due May 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: 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 3 - 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 4 - 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 5 - Data Bias
Readings: Cleveland, W. S., & McGill, R. (1984). "Graphical perception: Theory, experimentation, and application to the development of graphical methods." Journal of the American Statistical Association, 79(387), 531-554.

Week 6 - Data Summary and Visualization
Readings: Wang, Tricia. "Why Big Data Needs Thick Data." Ethnography Matters, 2016.

Week 7 - Qualitative Research and Mixed Methods
Readings: Sarker, I.H. "Machine Learning: Algorithms, Real-World Applications and Research Directions." SN COMPUT. SCI. 2, 160 (2021). https://doi.org/10.1007/s42979-021-00592-x

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 - SPRING BREAK

Week 10 - 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 11 - Human-Centered Algorithmic Auditing
Readings: Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). "Local Interpretable Model-Agnostic Explanations: An Introduction." O'Reilly.

Week 12 - Model Interpretability
Readings: Cifor, M., Garcia, P., Cowan, T.L., Rault, J., Sutherland, T., Chan, A., Rode, J., Hoffmann, A.L., Salehi, N., Nakamura, L. (2019). "Feminist Data Manifest-No." Retrieved from: https://www.manifestno.com/.

Week 13 - Data Feminism
Readings: Haerder, T; Reuter, A. (1983). "Principles of transaction-oriented database recovery." *ACM Computing Surveys*. 15(4): 287. doi:10.1145/289.291

Week 14 - Database Design
Readings: Tan Qiaoyu, Liu Ninhgao, and Hu Xia. (2019). "Deep Representation Learning for Social Network Analysis." *Frontiers in Big Data*. 2:2. doi:10.3389/fdata.2019.00002

Week 15 - Social Network Analysis
Readings: None

Week 16 - Careers in Human-Centered Data Science
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.

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.