

I310D - Introduction to Human-Centered Data Science

Monday/Wednesday: 2:00PM-3:30PM, WAG 201

Instructor: [Dr. Abhijit Mishra](#) (he/his)

Email: abhijitmishra@utexas.edu

TA: TBA

Email: TBA

Canvas: <https://utexas.instructure.com/courses/1352178>

Office Hours

Abhijit Mishra: Thursday 10:00AM-11:30AM via zoom (<https://utexas.zoom.us/j/8979599959>)

TA: TBA

Acknowledgement

I would like to thank [Maggie Engler](#), the previous instructor for I310D for generously sharing the course material and syllabus with me. The current syllabus is largely based on [Maggie's I310D Fall2022 syllabus](#).

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 abhijitmishra@utexas.edu and <TA>

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/co-requisite 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. Classes will be a mixture of lecture and hands-on sessions.

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. Additional materials/resources may be added to canvas prior to each class.

Required Devices

This course requires students to bring their laptop computers, although it is device agnostic. Students will be required to install Python, SQL and Jupyter notebooks.

Major Assignments

Reading Reflections (due weekly on Mondays 11:59AM unless specified otherwise)

Assignment 1 - Data Collection and Curation (due February 15 11:59 AM)

Assignment 2 - Data Bias (due March 22 11: 59AM)

Final Project (plan due April 12; presentation due April 29; report due May 1)

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%
Two 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-	60%
F	< 60%

Course Outline

All instructions, assignments, readings, rubrics and essential information will be on the Canvas website. Check the site regularly and use it to ask questions about the course schedule. Changes to the schedule may be made at my discretion and if circumstances require. For example, we might want to slow down, speed up or drop certain topics depending on student input. It is your responsibility to note these changes when announced.

WEEK 1. Introduction to Human-Centered Data Science

Lecture 1: Introduction, Course Syllabus, Other Concentration Courses, Career

Lecture 2: Data Science Problems and Solutions, Algorithmic thinking

References:

1. Provost, Foster, and Tom Fawcett. (2013). [Data science and its relationship to big data and data-driven decision making](#). *Big Data* 1.1 (2013): 51-59
2. Cao, L. (2017). [Data science: a comprehensive overview](#). *ACM Computing Surveys (CSUR)*, 50(3), 1-42.

WEEK 2. Ethics, Privacy and Consent + Intro to Programming for Data Science

Lectures: Research ethics and regulation in the USA, Consent and Contextual Integrity, Data legislation, Personal Identifiable Information, Differential Privacy

Hands-on: Data structure and programming basics, Setting up devices

References:

1. S. Barocas and H. Nissenbaum (2014). [Big Data's End Run Around Consent and Anonymity](#). In *Privacy, Big Data and the Public Good*. Cambridge University Press
2. Lim, Yish. (2019). [Doing Data the Right Way](#). Medium, Towards Data Science, 9 Jan. 2019, <https://towardsdatascience.com/doing-data-the-right-way-dbf5ba658177>.

WEEK 3. Research and Reproducibility + Open Source Software Development

Lectures: Introduction to Open Research, Reproducibility and Replicability, Best Practices in Research, OSS Development Philosophy, Data and Code Licenses

Hands-on: Project Collaborations using git

References:

1. Kitzes, J., Turek, D., & Deniz, F. (Eds.). (2018). The practice of reproducible research: case studies and lessons from the data-intensive sciences. *Univ of California Press* [Chapter 2: Assessing Reproducibility](#) and [Chapter 3: The Basic Reproducible Workflow](#)
2. Gienow, Michelle. (2018). [Tutorial: Git for Absolutely Everyone](#). <https://thenewstack.io/tutorial-git-for-absolutely-everyone/>

WEEK 4. Data Engineering: Data collection, Crowdsourcing

Lecture 1: Data engineering lifecycle, Data collection pipelines

Lecture 2: Resource generation strategies and crowdsourcing

References:

1. 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.
2. Amos, David. 2022. [A Practical Introduction to Web Scraping in Python](https://realpython.com/python-web-scraping-practical-introduction/). <https://realpython.com/python-web-scraping-practical-introduction/>

Assignment 1: Data Curation

WEEK 5. Data Engineering: Storytelling, Visualization

Lectures: Overview of Descriptive Statistics, Data visualization through plots and charts, Libraries and tools, Visualization rules

Hands-on: Data-frames, Matplotlib based plotting

References:

1. 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.
2. Python matplotlib tutorial by Jay Parmar: <https://blog.quantinsti.com/python-matplotlib-tutorial/>

WEEK 6. Database Design: Introduction to SQL, NoSQL

Lectures: Databases and types, Relational Databases, SQL Intro and examples, What is NoSQL, Difference between SQL and NoSQL

Hands-on: PostgreSQL setup and Basic SQL Query execution

References:

1. Anderson, B. (2022, June 12). [SQL vs. NoSQL databases: What's the difference?](#) IBM. Retrieved January 7, 2023, from <https://www.ibm.com/cloud/blog/sql-vs-nosql>

WEEK 7. Machine Learning: Introduction to ML and Deep Learning

Lecture 1: What is machine learning, Type of learners, Classification and Regression, ML Algorithms

Lecture 2: Neural Nets, Back propagation, Convolutional and Recurrent Neural Nets

References:

1. Raschka, S. (2020, August 5). [Chapter 1: Introduction to Machine Learning and Deep Learning](#). Sebastian Raschka, PhD. Retrieved January 7, 2023, from <https://sebastianraschka.com/blog/2020/intro-to-dl-ch01.html>

WEEK 8. Machine Learning: Unsupervised methods

Lecture 1: Intro to unsupervised methods, clustering, k-means algorithm,

Lecture 2: Anomaly detection, semi-supervised learning and transfer learning

References:

1. K-means clustering blog: <https://neptune.ai/blog/k-means-clustering>
2. Sarkar, D. (D. J. (2018, November 17). [A comprehensive hands-on guide to transfer learning with real-world applications in Deep learning](https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a). Medium. Retrieved January 7, 2023, from <https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a>

WEEK 9. Bias, Fairness, and Transparency

Lectures: Data bias, Algorithmic bias, Fairness and Accountability

Hands-on: Data bias example using python

References:

1. Ingold, David and Soper, Spencer. (2016). [Amazon Doesn't Consider the Race of Its Customers. Should It?](#). *Bloomberg*
2. Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner. (2018). [Machine Bias: Risk Assessment in Criminal Sentencing](#). *ProPublica*, May 2018.

Assignment 2: Data Bias, Project group formation

WEEK 10. SPRING BREAK

WEEK 11. Model Interpretability and Explainability

Lectures: Algorithmic Auditing, Importance of Interpreting, Interpretability vs Explainability, LIME, Interpretability in Neural Nets

Project: Specification and problem statement given

References:

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

WEEK 12. Qualitative Research and Mixed Methods

Lectures: Qualitative research methods, Ethnography, big-vs-thick data, goodness of fit and data requirements, metrics

References:

1. Wang, Tricia. [Why Big Data Needs Thick Data](#). *Ethnography Matters*, 2016.

WEEK 13. Human-Centered Social Network Analysis

Lectures: Social Network Platforms, Foundations and example of Social Network Analysis, Ethical aspects

Hands-on: Sentiment analysis on twitter data using existing NLP models

References:

1. Extracting Data from Twitter using Python: <https://towardsdatascience.com/extracting-data-from-twitter-using-python-5ab67bff553a>

WEEK 14: Collaborations across and beyond Data Science

Lectures: Working in organizations and communities, Working across disciplines, Handling people's data, Human AI collaborations

Project: Planning Intermediate Progress Discussion

Readings: No readings

WEEK 15. Intellectual Properties: Publishing and Patenting + Data Science Career Options

Lectures: How to read data-science papers effectively, Protecting IPs, Publications vs Patenting, Patenting process

Readings: No readings

WEEK 16. Project Presentations

WEEK 17. NO FINAL EXAM

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.

Title IX Reporting

Title IX is a federal law that protects against sex and gender-based discrimination, sexual harassment, sexual assault, unprofessional or inappropriate conduct of a sexual nature, dating/domestic violence and stalking at federally funded educational institutions. UT Austin is committed to fostering a learning and working environment free from discrimination in all its forms. When unprofessional or inappropriate conduct of a sexual nature occurs in our community, the university can:

1. Intervene to prevent harmful behavior from continuing or escalating.
2. Provide support and remedies to students and employees who have experienced harm or have become involved in a Title IX investigation.
3. Investigate and discipline violations of the university's relevant policies.

Beginning January 1, 2020, Texas Senate Bill 212 requires all employees of Texas universities, including faculty, report any information to the Title IX Office regarding sexual harassment, sexual assault, dating violence and stalking that is disclosed to them. Texas law requires that all employees who witness or receive any information of this type (including, but not limited to, writing assignments, class discussions, or one-on-one conversations) must be reported. **I am a Responsible Employee and must report any Title IX related incidents** that are disclosed in writing, discussion, or one-on-one. Before talking with me, or with any faculty or staff member about a Title IX related incident, be sure to ask whether they are a responsible employee. If you would like to speak with someone who can provide support or remedies without making an official report to the university, please email advocate@austin.utexas.edu. For more information about reporting options and resources, visit <http://www.titleix.utexas.edu/>, contact the Title IX Office via email at titleix@austin.utexas.edu, or call 512-471-0419.

Although graduate teaching and research assistants are not subject to Texas Senate Bill 212, they are still mandatory reporters under Federal Title IX laws and are required to report a wide range of behaviors we refer to as unprofessional or inappropriate conduct of a sexual nature, including the types of conduct covered under Texas Senate Bill 212. The Title IX office has developed supportive ways to respond to a survivor and compiled campus resources to support survivors.