

# I310D - Introduction to Human-Centered Data Science

**Tuesdays / Thursdays : 3:30PM-5:00PM, PAR 203**

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

**Email:** [abhijitmishra@utexas.edu](mailto:abhijitmishra@utexas.edu)

**TA:** Krishna Sri Somepalli (she/her)

**Email:** [krishnasri@utexas.edu](mailto:krishnasri@utexas.edu)

**Canvas:** <https://utexas.instructure.com/courses/1367918>

## Office Hours

Abhijit Mishra: Thursday 8:00AM-10:00AM 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](mailto:abhijitmishra@utexas.edu) and Krishna ([krishnasri@utexas.edu](mailto:krishnasri@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/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. As of now, there are **no plans to record lectures**.

### **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.

### **Assignments and Project**

- 1. Written Assignments: FIVE** Reading Reflections (Due on Mondays in the first half of the semester)
- 2. Coding Assignment 1** - Data Collection and Curation
- 3. Coding Assignment 2:** Database operations
- 4. Coding Assignment 3** - Data Bias
- 5. Course Project (group-based)**

There are no traditional examinations within this course. Instead, your assessment will encompass written assignments, coding projects, and collaborative group assignments.

### **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.

### **I310D 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.

<b>Grade Component</b>	<b>Percentage</b>
Class participation and attendance (i.e., all graded hands on exercises)	20%
Reading reflections	15%
Final project	35%
Three Assignments	30%

### **Grade Breaks**

<b>Grade</b>	<b>Cutoff</b>
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 (08/22/2023):** Introduction, Course Syllabus, Other Concentration Courses in HCDS

**Hands-on (08/24/2023):** Basic python programming for Data Science 1, Setting up devices

#### 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.
3. Introduction to Python: <https://www.datacamp.com/courses/intro-to-python-for-data-science>
4. [Virtual environment + Jupiter] <https://towardsdatascience.com/python-virtual-environments-jupyter-notebook-bb5820d11da8> Links to an external site.

**Assignment: Reading Reflection 1 (Posted in Canvas)**

### WEEK 2. Data Science : Approaches, Tasks, Solutions, Algorithmic Thinking

**Lecture (08/29/2023):** Data Science Problems and Solutions, Algorithmic thinking

**Hands-on (08/31/2023):** More on python programming, Implementing algorithms

#### References:

1. A beginner's guide to algorithmic thinking: <https://learntocodewith.me/posts/algorithmic-thinking>
2. Python Data Structures: <https://www.geeksforgeeks.org/python-data-structures/>

**Assignment: Reading Reflection 2 (Posted in Canvas)**

### WEEK 3. Ethics, Privacy and Consent

**Lecture (09/05/2023):** Privacy and Fair Information Practices, Sensitive and Private Data, Data Anonymization Techniques, Differential Privacy, Data legislation,,

**Hands-on (09/07/2023):** Python advanced data types, libraries and loading data in Pandas

**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>
3. Pandas Tutorial: <https://www.w3schools.com/python/pandas/default.asp> (Section: Basic)

**Assignment: Reading Reflection 3 (Posted in Canvas)**

**WEEK 4. Research and Reproducibility + Open Source Project Development**

**Lectures (09/12/2023):** Introduction to Open Research, Reproducibility and Replicability, Best Practices in Research, OSS Development Philosophy, Data and Code Licenses

**Hands-on (09/14/2023):** Git basics and project creation and collaboration 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/>

**Assignment: Reading Reflection 4 (Posted in Canvas)**

**WEEK 5. Data Engineering I: Data collection, Crowdsourcing**

**Lecture 1 (09/19/2023):** Data engineering lifecycle, Data collection pipelines, Resource generation strategies and crowdsourcing

**Hands-on (09/21/2023):** Web-scraping using Python

**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/>

**Assignment: Reading Reflection 5 (Posted in Canvas)**

## WEEK 6. Data Engineering II: Storytelling, Visualization

**Lecture (09/26/2023):** Overview of Descriptive Statistics, Data visualization through plots and charts, Libraries and tools, Visualization rules

**Hands-on (09/28/2023):** 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/>

## Programming Assignment 1: Data Curation (Posted in Canvas)

## WEEK 7. Database Design I: Introduction to SQL, NoSQL

**Lecture (10/03/2023):** Databases and types, Relational Databases, SQL Intro and examples, What is NoSQL, Difference between SQL and NoSQL

**Hands-on (10/05/2023):** 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 8. Database Design II: More on SQL and Database Operations

**Lecture (10/10/2023):** Selecting data through SELECT statement and WHERE clause, Data Definition Language (DDL) - CREATE, ALTER and DROP Tables, Data Manipulation Language (INSERT, UPDATE, DELETE statements)

**Hands-on (10/12/2023):** PostgreSQL setup and Basic SQL Query execution

### References:

1. Handout DDL and DML : <https://www.eecs.yorku.ca/~papagel/courses/eecs3421/docs/lectures/05-dml-ddl-views-indexes.pdf>
2. Handout SQL DML : [https://course.ccs.neu.edu/cs520of17/ssl/lectures/lecture\\_03\\_sql\\_1.pdf](https://course.ccs.neu.edu/cs520of17/ssl/lectures/lecture_03_sql_1.pdf)
3. SQL Basics and Joins: [https://ocw.mit.edu/courses/1-204-computer-algorithms-in-systems-engineering-spring-2010/ca8f491f5f7f57335b87a52700e54849/MIT1\\_204S10\\_lec03.pdf](https://ocw.mit.edu/courses/1-204-computer-algorithms-in-systems-engineering-spring-2010/ca8f491f5f7f57335b87a52700e54849/MIT1_204S10_lec03.pdf)

## Programming Assignment 2: Database Operations (Posted in Canvas)

## WEEK 9. Machine Learning I: Introduction to ML and Deep Learning

**Lecture 1 (10/17/2023):** What is machine learning, Type of learners, Classification and Regression, ML Algorithms, Brief Introduction to Neural Network and Deep Learning

**Lecture 2 (10/19/2023):**, Hands-on: ML on classification data

### References:

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

## WEEK 10. Machine Learning II: Unsupervised methods

**Lecture (10/24/2023):** Intro to unsupervised methods, clustering, k-means algorithm, Anomaly detection,

**Hands-on (10/26/2023):** Anomaly detection on time-series data

### 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>

**Project: Proposal and Planning (Posted in Canvas)**

## WEEK 11. Bias, Fairness, and Transparency

**Lecture (10/31/2023):** Data bias, Algorithmic bias, Fairness and Accountability

**Hands-on (11/02/2023):** 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.

**Programming Assignment 3: Data Bias (Posted in Canvas)**

## **WEEK 12. Algorithmic Auditing, Model Interpretability, Qualitative Research, Mixed Methods**

**Lecture 1 (11/07/2023):** Algorithmic Auditing, Importance of Interpreting, Interpretability vs Explainability, LIME, Interpretability in Neural Nets

**Lecture 2 (11/09/2023):** Qualitative research methods, Ethnography, big-vs-thick data, goodness of fit and data requirements, metrics

### **References:**

1. Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). [Local Interpretable Model-Agnostic Explanations: An Introduction](#). O'Reilly.
2. Wang, Tricia. [Why Big Data Needs Thick Data](#). Ethnography Matters, 2016.

## **WEEK 13. Collaborations Across and Beyond Data Science + 3 Minute Project Progress Presentation**

**Lecture 1 (11/14/2023):** Working in organizations and communities, Working across disciplines, Handling people's data, Human AI collaborations

**In-class presentation 2 (11/16/2023):** Project Progress Presentation by groups and feedback soliciting

### **References:**

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

## **WEEK 14: Fall break / Thanksgiving; no classes held**

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

**Lecture (11/28/2023):** How to read data-science papers effectively, Protecting IPs, Publications vs Patenting, Data Science Career Options

**Lecture (11/30/2023):** FINAL Group Project PRESENTATION (2 hours class)

## **WEEK 16. STUDY DAYS**

FINAL Report Submission by December 6 (12/06/2023)

## **WEEK 17. NO FINAL EXAM**

Final grades submission by instructor by December 10 (12/10/2023)

## **Mantra for student success : Navigating the I310D HCDS course**

- Achieve higher attendance, aiming for 100% to maximize exposure and engagement during lectures and practical exercises.



- Submit practicums and assignments promptly, recognizing that minor errors can be overlooked while focusing on continuous improvement.
- Prioritize transparency by appropriately citing tools, resources, and data sources, showcasing your commitment to ethical and accountable work.
- Approach in-class quizzes with a clear understanding and well-organized thoughts, leveraging your conceptual clarity to excel.
- If programming presents challenges, embrace deliberate practice to strengthen your skills and confidently navigate technical aspects. Repeatedly and extensively seek help from the instructor / TA during office hours.
- Embrace iteration as you prepare presentations, ensuring impactful task demonstrations, comprehensive analyses, and well-structured reports.
- Recognize that success in the course is a result of these concerted efforts, culminating in your growth as a proficient and accomplished data practitioner.

### 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>.

### AI Tools Usage Policy:

The utilization of AI-powered tools, including platforms like ChatGPT, DALL-E, or any other large language/image generative models, to create content such as text, code, images, multimedia, or any related materials intended for assignments, quizzes, or projects that contribute directly to the evaluation of grades within this course is **strictly proscribed**. Exceptions to this rule apply only if the incorporation of such systems aligns with the specified objectives of the assignment or project. Breaching this policy may result in the initiation of proceedings related to student misconduct.

Should there be any suspicion surrounding the content submitted by a student, suggesting the involvement of an AI tool, I retain the authority to request clarification from the student. This clarification may be sought through email communication or arranged verbal discussions in the form of one-on-one meetings. In the event of any inconsistencies between the provided explanations and the submitted solutions, I reserve the right to instigate misconduct proceedings against the concerned student. Upon enrolling in this course, students inherently express their agreement to adhere to this policy as well as any forthcoming policies described below.

### Course Material Sharing Policy

Unauthorized sharing or distribution of lecture notes, slides, or examination questions is strictly prohibited without prior permission from the instructors. Failure to adhere to this policy may result in the initiation of legal actions. 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 these restrictions 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](mailto: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](mailto: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.