Data and Society, Spring 2026

Class info

I 320S / 320J / 320C (28607, 28644, 28665)
GAR 3.116
MW 2:00 - 3:30pm.

Instructor

Hanlin Li, PhD, she/her, Assistant Professor
Email: lihanlin@utexas.edu Office: UTA 5.412
Office hour: TBD and by appointment

Course Description

What role does data play in our society? What infrastructure and systems does data traverse? And what technologies and practices are associated with data? Who does an increasingly data-driven society affect the most? In this course, students will answer these questions through discussion, readings, and projects. Students will explore common data collection, management, and sharing practices around information technology and emerging technologies such as AI. Students will gain hands-on experiences with collecting and managing user data in ethical and responsible manners. Students will design data-driven systems that are centered around user consent, transparency, and social responsibilities.

Long Description

What role does data play in our society? What infrastructure and systems does data traverse? And what technologies and practices are associated with data? Who does an increasingly data-driven society affect the most? In this course, students will answer these questions through discussion, readings, and projects.

This course explores the intersection of data collection, technology, and society, focusing on the ethical, social, and legal issues that arise from large-scale data collection practices. Students will critically examine how data is gathered, analyzed, and used by governments, corporations, and other institutions, and its impact on privacy, equity, and power dynamics. Students will delve into topics such as data work, crowdsourcing, surveillance, privacy, algorithmic bias, informed consent, data justice, and the implications of emerging technologies like AI and IoT. The course also emphasizes the role of policy, activism, and design in shaping a more equitable and ethical data landscape.

By the end of the course, students will be equipped to evaluate and critique data practices, understand their societal consequences, and propose solutions that prioritize ethical considerations and social good. This course is ideal for those interested in the fields of data science, public policy, technology ethics, and social justice.

Learning outcomes

Students will learn how to examine what role data plays in our increasingly data-driven society and the pros and cons of different data collection, management, and sharing practices through readings, discussions, and case studies.

Students will gain hands-on experience with data management or systems as UX designers, researchers, and data scientists by completing a class project.

Students will also be exposed to interdisciplinary research on the impact of data-driven technologies and important ethical considerations, e.g. privacy and consent.

How will you learn
This course uses a blended strategy of student-led discussions, mini-lectures, and asynchronous assignments. In addition to attending and participating in discussions and lectures, students will be expected to contribute to Canvas discussion and complete a semester-long design project.

Classroom norms and participation
This class will be centered around social learning. This means that the instructor will foster a constructive, inclusive discussion space for all students and students are expected to attend all class sessions in person and participate actively in class discussion. Class participation will be graded.

Students should only use laptops for note taking purposes. The instructor may limit technology use to minimize classroom distractions.

If you are a student with a disability, or think you may have a disability, and need accommodations please contact Disability and Access (D&A). You may refer to D&A's website for contact and more information: http://disability.utexas.edu/. If you are already registered with D&A, please deliver your Accommodation Letter to me as early as possible in the semester so we can discuss your approved accommodations.

How will you be evaluated
Students will be evaluated on their completion of assignment and their ability to apply knowledge to several short-term assignments and their design project.

Major assignments
Reading responses every week - 32%
Students will complete reading assignments each week for ten weeks and submit their reading responses on Canvas the day before class. Each reading response will account for 3%, except Week 1's (2%). These reading responses are meant to help students reflect on what they have learned, what they disagree with the readings, and what lingering questions they have. Reading responses should be of reasonable length (200+ words). Your reading responses will be graded for substance and originality. I expect you would spend the first two sentences summarizing the readings and the rest of the response would be to elaborate your reflections and questions. You

can relate it to your personal experiences, your own coursework or projects, or other courses you have taken. No late work will be accepted.

Individual assignment - Google crowdsource & Mozilla Common Voice 8%
Students will complete two crowdsourcing tasks, one on Google Crowdsource and one on Mozilla Common Voice. These two platforms' purpose is to collect data from volunteers for AI model training. You will perform tasks such as image labeling, audio transcription, and audio recording on these two platforms. You will then write up a report working on the two platforms. Your writeup should be 1500 words long and you can take screenshots of these two platforms to support your argument. I will provide suggested questions for you to answer in your writeup.

Individual assignment - datasheet 8%
Students will produce a datasheet to accentuate critical information about a given dataset. I will provide several datasets as options and you will choose one and read through its existing documentations. You will then perform necessary descriptive analysis and distill critical information about the dataset into a formatted datasheet. You will also include a reflection in your assignment, in response to suggested questions.

Design project - 32%
Students will complete a semester-long design project, focusing on responsible data collection. The project will have multiple milestones, including a project proposal, a midterm presentation, peer reviews, and a final presentation. Detailed instructions will be provided for each milestone.

Participation - 20%
Students will lead in-class discussion once during the semester (10%) and complete activities during class (10%). Attendance will be tracked each week and the lowest attendance grade will be dropped.

Late submission policy
Late reading responses and late presentations will not be accepted.

For other assignments, instead of asking for extensions, you are given seven "slip day" credits for the semester; you will be charged a "slip day" for each day an assignment is late. A slip day is accrued starting immediately after the assignment is due (i.e. an assignment which is one hour late will incur a full slip day). This means you can hand in a submission up to seven days after the due date. You can also distribute your slip day credits across multiple assignments, e.g. one assignment 3 days later and the other 4 days late, without penalty. Once all of your slip day credits have been applied, any late submission will result in 5% grade deduction per day.

Required and Recommended class materials
Readings will include news reports, academic publications, and book chapters. All class materials will be provided via Canvas.

Weekly Schedule

This schedule is subject to change, pending guest speaker availability. I will announce any change two weeks in advance.

3		Crowdsourcing	https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html
	Jan 29		Kotaro Hara, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P. Bigham. 2018. A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). Association for Computing Machinery, New York, NY, USA, Paper 449, 1–14. https://doi.org/10.1145/3173574.3174023Links to an external site.
4		data work	Roberts, Sarah T.. "2. Understanding Commercial Content Moderation". Behind the Screen: Content Moderation in the Shadows of Social Media, New Haven: Yale University Press, 2019, pp. 33-72. https://doi.org/10.12987/9780300245318-003
	Feb 5		Who does the work of Data;
5		scraping and passive data collection	This is where the data to build AI comes from
	Feb 12		Exploring 12 million of the images used to train Stable Diffusion
6		consent and privacy	AI Consent Futures: A Case Study on Voice Data Collection with Clinicians
	Feb 19		what's next for our privacy - mit tech review [uploaded to canvas]
7		documentation	Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets;
	Feb 26		Bandy, J., & Vincent, N. (2021, June). Addressing" documentation debt" in machine learning: A retrospective datasheet for bookcorpus. In Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1).
8		Research ethics	Enhancing the ethics of user-sourced online data collection and sharing;
	Mar 3		"Participant" Perceptions of Twitter Research Ethics. Casey Fiesler and Nicholas Proferes
9		mid term presentation	
10		[Spring break]	
11	Mar 24	[data literacy activity]	

12		subjectivity and bias	A Taxonomy of Challenges to Curating Fair Datasets;
	Apr 2		why algorithms can be racist and sexist:
13		Sharing	We Research Misinformation on Facebook. It Just Disabled Our Accounts;
	Apr 9		A Systematic Review of NeurIPS Dataset Management Practices
14		Scientific data	Data sharing practices and data availability upon request differ across scientific disciplines
	Apr 16		https://data.research.cornell.edu/data-management/sharing/fair/
15	Apr 21	Final presentations	
	Apr 23	Final presentations	
16	Apr 28	drop in discussion and next steps	