Harvard Summer School
HARVARD DIVISION OF CONTINUING EDUCATION

STAT S-115
Data Science: An Artificial Ecosystem

Summer 2025

Course Information

CRN: 35611
Section Number: 1
Format: Live Attendance Web Conference
Credit Status: Undergraduate, Graduate, Noncredit
Credit Hours: 4

Course Description: This course aims to introduce students to the world of the data science via articles published in the *Harvard Data Science Review*, a global forum disseminating everything data science and data science for everyone. The course emphasizes the evolutionary nature of the data science enterprise as an artificial ecosystem, where the phrase artificial shares a similar connotation as it is in the phrase artificial intelligence (AI). However, unlike the common algorithmic or robotic depictions of AI, this course espouses a panoramic view of data science, from philosophical conceptualization of data to interpretation and policy implications of statistical findings and to the re-use of data for addressing scientific replicability and reliability. Topics such as generative AI, data generation, and data privacy are explored from multiple perspectives to demonstrate the necessity of the panoramic approach. Questions such as what is intelligence or what is data require philosophical contemplation, while assessing the positive and negative impact of generative AI or digital technologies in general demand careful sociological, computational, and statistical thinking. Furthermore, determining how to ensure effective and safe human-computer interaction requires advanced data science theory and methods. Throughout the course, students engage with and critique a broad range of data science articles that incorporate perspectives from

computer science, statistics, philosophy, social sciences, and other fields of study. During course meetings, there are many opportunities to discuss some of these articles with the authors themselves.

Instructor Information & Office Hours

Xiao-li Meng
Email: xlmeng@fas.harvard.edu

Course Goals / Learning Outcomes

UPDATED COURSE DESCRIPTION for 2025 Summer

This course aims to introduce students to the world of data science and artificial intelligence (AI) technologies via articles published in the Harvard Data Science Review, a global forum disseminating everything data science and data science for everyone. The course emphasizes the evolutionary nature of the data science enterprise as an artificial ecosystem, where the term “artificial” shares a similar connotation as it does in the phrase “artificial intelligence”. However, unlike the common algorithmic or robotic depictions of AI, this course espouses a panoramic view of data science and AI technologies, from philosophical conceptualizations to interpretation and policy implications of data science and AI technologies. Topics such as generative AI and human-machine interaction are explored in depth to demonstrate the necessity of the panoramic approach. Questions such as “What is intelligence?” or “How might AI impact humans as emotional beings?” require philosophical contemplation, while assessing the positive and negative impacts of generative AI demand careful sociological, economical, computational, and statistical thinking. Furthermore, determining how to ensure the reliability of AI technologies as well as effective and safe human-machine interaction requires advanced theory and methods. Throughout the course, students engage with and critique a broad range of articles by leading scholars and practitioners with perspectives from computer science, economics, statistics, philosophy, sociology, and more. During course meetings, there are ample opportunities to discuss these articles with the authors themselves, as well as to investigate the use of generative AI for enhancing learning experiences.

Prerequisites: Students are not required to have any prior experience in computer science, statistics, or data science. Background tutorials will be provided as necessary. The course

assignments will assess students' ability to think critically and communicate effectively, rather than their computational or mathematical skills.

Course Goals

Throughout the course, students will engage with and critique a broad range of articles on data science and AI, drawing perspectives from computer science, statistics, philosophy, social sciences, and more. During the course meetings, there will be opportunities to discuss many articles with their original authors. The ultimate goal of the course is to introduce students to the AI and data science ecosystems, which are far broader and more complex than the commonly perceived data analytic world that focuses on data processing and analysis.

Learning Outcomes

1. Interdisciplinary Appreciation and Integration:
    - Students will identify and articulate the key tenets and perspectives of data science and AI technologies as presented in the Harvard Data Science Review articles, encompassing fields such as computer science, statistics, philosophy, and social sciences.
    - Students will synthesize and integrate knowledge across these disciplines, demonstrating a solid understanding of data science and AI technologies as holistic enterprises rather than a collection of isolated technical tasks and algorithms.
2. Critical Thinking and Evaluation:
    - Students will develop the ability to critically assess and evaluate the strengths, weaknesses, and broader implications of the articles studied.
    - They will be able to pose critical questions and engage in informed discussions, showcasing their ability to understand, critique, and extend the ideas presented in the readings.
3. Ethical and Philosophical Reflection:
    - By engaging with questions such as “What is intelligence?” or “What can and should AI do?”, students will cultivate the ability to critically examine and

articulate the philosophical and ethical dimensions of data science and AI technologies.

- They will demonstrate a nuanced understanding of complex topics like generative AI and human-machine interaction, and appreciate the interplay between technological advancements and societal implications

These outcomes capture the essence of the course which lies at the intersection of technical depth and philosophical breadth, emphasizing a comprehensive understanding of data science and AI technologies.

Mode of Attendance & Participation Policy

Class meetings take place over Zoom. Because they involve active participation, discussion, and dialogue, you are expected to attend all class meetings. Please arrive on time. You should attend Zoom meetings with a functional web-camera and microphone, prepared with materials needed, to engage thoughtfully, and with your camera on. You may turn off your camera for occasional interruptions or momentarily for privacy.

You will also need the most up-to-date Zoom client installed on your computer to join class. Please participate from a safe and appropriate environment with appropriate clothing for class. Participating while traveling or in a car is not permitted. In addition, please do not join class via mobile phone or web browser.

Please be sure to review important information on **Student Responsibilities and Policies**.

Assignments & Grading

Course Requirements and Assignments

The primary deliverables in the course are two essays and corresponding presentations:

- For each essay, students will submit a first draft of 1,000 words maximum and obtain feedback prior to submitting a final draft that is 500 words maximum.
- Students will give a short presentation of each essay during class time. Tentatively, each presentation should be at most 8 minutes; this is subject to change depending on how many students enroll in the course. The presentations will take place after submission of the first draft and prior to submitting the final draft so that students

can incorporate peer feedback as well as instructor feedback into their essay, in addition to gaining valuable experience communicating their ideas verbally.

Students are expected to follow the course reading list and actively participate in class discussions. To support this, there will be short pre-class assignments designed to help students engage critically with the readings and prepare thoughtful questions for guest speakers. After each class, students will complete post-class assignments to encourage reflection on the discussion. These assignments will be submitted on Gradescope and shared as discussion posts in Slack.

Further details about the assignments are available here.

Grading

1. Participation in class as well as via the pre/post-class assignments (40%)
2. Two essays (15% per essay, 30% in total)
3. Two class presentations (15% per presentation, 30% in total)

Grade Definitions

Students registered for undergraduate or graduate credit who complete the requirements of a course may earn one of the following grades:

A and A– Earned by work whose superior quality indicates a full mastery of the subject—and in the case of A, work of extraordinary distinction. There is no grade of A+.

B+, B, and B– Earned by work that indicates a strong comprehension of the course material, a good command of the skills needed to work with the course materials, and the student's full engagement with the course requirements and activities.

C+, C, and C– Earned by work that indicates an adequate and satisfactory comprehension of the course material and the skills needed to work with the course materials, and that indicates that the student has met the basic requirements for completing assigned work and participating in class activities.

D+, D, and D– Earned by work that is unsatisfactory but that indicates some minimal command of the course materials and some minimal participation in class activities that

is worthy of course credit.

E Earned by work that is unsatisfactory and unworthy of course credit. This grade may also be assigned to students who do not submit required work in courses from which they have not officially withdrawn by the withdrawal deadline. Zero or E grades are assigned to students for missing work. These grades are included in the calculation of the final grade.

Graduate Credit Requirements

For students taking the course for graduate credit, they will need to read an additional research article from HDSR (that has been approved by the instructor) and write a "Media Summary" for the article that is no more than 400 words (and does not already have a media summary); a media summary is a plain-language summary of an article for general audiences and press releases.

Reading List

This semester, we will be exploring two themes: 1) how we can use GAI to solve problems, 2) how we can solve problems posed by GAI. Subject to author availability, we plan on discussing the following articles.

Theme 1: How can we use GAI to solve problems?

- Data Science and AI in Context: Summary and Insights (Alfred Spector)
- A Unified Framework of Five Principles for AI in Society (Luciano Floridi and Josh Cowls)
- Can ChatGPT Plan Your Retirement?: Generative AI and Financial Advice (Andrew Lo and Jillian Ross)
- Confidence in the Reasoning of Large Language Models (Yudi Pawitan and Chris Holmes)
- An Information-Theoretic Approach for Detecting Edits in AI-Generated Text (Idan Kashtan and Alon Kipnis)
- Demonstrations of the Potential of AI-based Political Issue Polling (Nathan E. Sanders, Alex Ulinich, and Bruce Schneier)

- Effective Generative AI: The Human-Algorithm Centaur (Soroush Saghafian and Lihi Idan)

Theme 2: How can we solve problems posed by GAI?

- One Person Dialogues: Concerns About LLM-Human Interactions (Darren Frey and Daniel H. Weiss)
- AI, Language, and the Humanities (Gregory Crane)
- Toward a Theory of AI Errors: Making Sense of Hallucinations, Catastrophic Failures, and the Fallacy of Generative AI (Veronica Barassi)
- Beware the Intention Economy: Collection and Commodification of Intent via Large Language Models (Yaqub Chaudhary and Jonnie Penn)
- Voices in the Code (available for free via the Harvard Library) and Participatory Engineering of Algorithmic Social Technologies: An Extended Book Review of David G. Robinson's Voices in the Code (Rajesh Venkatachalapathy and David G. Robinson)
- How to Define and Execute Your Data and AI Strategy (Ulla Kruhse-Lehtonen and Dirk Hofmann)
- AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap (Q. Vera Liao and Jennifer Wortman Vaughan)

Community Charter

By enrolling in courses offered by the Harvard University Division of Continuing Education (DCE), individuals agree to abide by our community standards to promote a culture of trust, cooperation, mutual understanding, and learning. Members of the DCE community, including students, faculty, and staff, are expected to treat each other with dignity and communicate respectfully and appropriately across all channels. Our commitment to academic integrity and excellence includes holding ourselves accountable for our actions. Violations of these community standards may result in disciplinary actions.

DCE adheres to all University-wide policies that address discrimination, bullying, and harassment. Resources including the Office for Gender Equity and the Harvard Ombuds Office are available to assist community members with concerns.

Academic Integrity Policy

You are responsible for understanding Harvard Summer School policies on Academic Integrity and how to use sources responsibly. Violations of academic integrity are taken very seriously. Visit Resources to Support Academic Integrity and the Harvard Guide to Using Sources to review important information on academic citation rules.

AI Technologies. The Summer School's Academic Integrity Policy. prohibits students from representing work as their own that they did not write, code, or create. It is never permissible to submit work generated by machine learning and AI technologies (such as ChatGPT) without proper attribution.

Course Policy in Using AI Technologies.

With the arrival of generative AI (GAI), there has been a heated debate about the use of such technology for assisting with coursework. We believe that you are taking this course to gain a deeper appreciation of what data science entails. With that in mind, we trust that you will exercise your best judgment in deciding when and how to take advantage of AI technologies to facilitate your learning.

As part of course assignments, we will ask you to experiment with using GAI and reflect on your experiences. If you use GAI in any other substantial way that was not explicitly part of the assignment instructions, be sure to include a note in your submission describing how you used it; consult the teaching team in advance if you are unsure whether your proposed use of GAI is acceptable. For example, it is permissible to use GAI for sentence-level editing (such as suggesting different synonyms, rephrasing text that you wrote, etc.) but all the reflective work in an essay that you submit must be your own.

Accessibility Services Policy

The Division of Continuing Education (DCE) is committed to providing an accessible academic community. The Accessibility Services Office (ASO) is responsible for providing accommodations to students with disabilities. Students must request accommodations or adjustments through the ASO. Instructors cannot grant accommodation requests without prior ASO approval. It is imperative to be in touch with the ASO as soon as possible to avoid delays in the provision of accommodation.

DCE takes student privacy seriously. Any medical documentation should be provided directly to the ASO if a substantial accommodation is required. If you miss class due to a short-term illness, notify your instructor and/or TA but do not include a doctor's note. Course staff will not request, accept, or review doctor's notes or other medical documentation. For more information, email accessibility@extension.harvard.edu.

Publishing or Distributing Course Materials Policy

Students may not post, publish, sell, or otherwise publicly distribute course materials without the written permission of the course instructor. Such materials include, but are not limited to, the following: lecture notes, lecture slides, video, or audio recordings, assignments, problem sets, examinations, other students' work, and answer keys. Students who sell, post, publish, or distribute course materials without written permission, whether for the purposes of soliciting answers or otherwise, may be subject to disciplinary action, up to and including requirement to withdraw. Further, students may not make video or audio recordings of class sessions for their own use without written permission of the instructor.

Canvas Access After End of Session

The Canvas website for this course will remain available to enrolled students for a limited time after the course concludes. **You are encouraged to download coursework and materials you wish to keep before the Summer session ends**. See Course Formats & Required Technology for additional information on Canvas access.

Class Meeting Schedule

Please note that the class meeting schedule is subject to change.

Week 1: June 23/25

- Class 1: Overview and Introduction
- Class 2: Tutorial on Probability, Inference, and Modeling

Week 2: June 30 / July 02

- Class 3: Guest Speaker(s) on Theme 1
- Class 4: Guest Speaker(s) on Theme 1

Week 3: July 07/09

- Class 5: Guest Speaker(s) on Theme 1 (Essay 1 first draft due)
- Class 6: AI-Assisted Learning Activity on Theme 1

Week 4: July 14/16

- Class 7: Student Presentations on Theme 1
- Class 8: Guest Speaker(s) on Theme 2 (Essay 1 final draft due)

Week 5: July 21/23

- Class 9: Guest Speaker(s) on Theme 2 (Essay 2 first draft due)
- Class 10: Student Presentations on Theme 2

Week 6: July 28/30

- Class 11: Guest Speaker(s) on Theme 2
- Class 12: Guest Speaker(s) on Theme 2

Week 7: Aug 04/06

- Class 13: Guest Speaker(s) on Theme 2 (Essay 2 final draft due)
- Class 14: Grand Finale Guest Speaker